End User Documentation **CUSTOMER**

Document Version: 1.0 - 2014-11

SAP InfiniteInsight® 7.0

Data Toolkit User Guide

Table of Contents

1 About this Document 3

2 Common Steps 4

2.1 Selecting a Data Set 4

2.1.1 Special Case: Data Stored in Databases - the Explain Mode 5

2.1.2 Sub-sampling a Data Set 6

2.2 Describing the Data Set 7

3 Data Viewer 9

3.1 Visualizing a Data Set 10

3.2 Deleting a Data Set 11

4 Data Transfer 12

4.1 Transferring a Data Manipulation 13

4.1.1 Mapping Fields 14

4.2 Selecting Variables 14

4.3 Saving the New Data Set 15

4.3.1 Defining Text File Settings 15

5 List Distinct Values 17

5.1 Selecting the Variable to Extract 18

5.2 Saving the Categories 18

6 Descriptive Statistics 20

6.1 From the Start Panel 20

6.1.1 Selecting the Variables 21

6.1.2 Checking the Parameters 24

6.1.3 Generating the Statistics 28

6.1.4 Statistics Debriefing 28

6.2 From the Panel "View Data" 29

6.2.1 Computing the Statistics 30

6.2.2 Viewing the Statistics 31

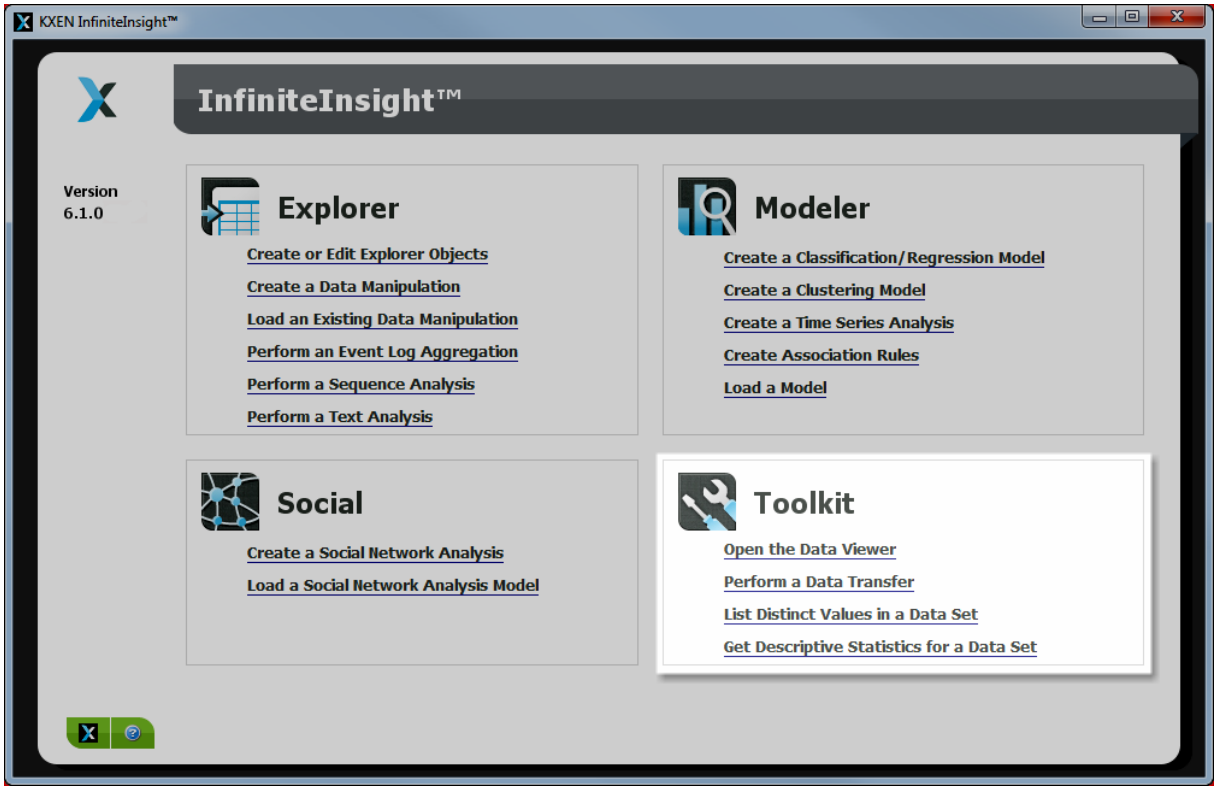
6.3 Geo-location 32

6.3.1 Using Geolocation Feature 33

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

1 About this Document

This document presents the **InfiniteInsight® Toolkit** provided in SAP InfiniteInsight®.



The features offered in this toolkit allow you to:

- explore and delete data sets (*Open the Data Viewer*).
- transfer a data source to another location or format (*Perform a Data Transfer*).
- export the list of distinct values of a variable from your data set. Among other things, this feature allows you to generate a list of unique index values from an existing data set and to use it as a reference table when performing an **event log aggregation** or a **sequence analysis** in InfiniteInsight® Explorer (*List Distinct Values in a Data Set*).
- generate variables statistics for a data set, such as the cross-statistics or the profit curves when a variable in the data set is used to estimate the target variable (*Get Descriptive Statistics for a Data Set*).

Before reading this document you should be familiar with the section Essential Concepts of the **InfiniteInsight® - User Guide**.

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

2 Common Steps

This section of the document describes:

- the data source selection,
- the data description

which are steps common to several of the Toolkit features.

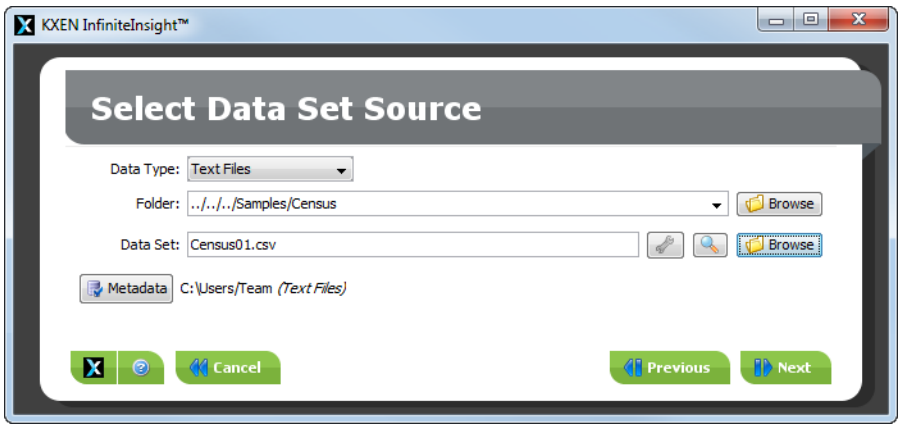
IN THIS CHAPTER

Selecting a Data Set..... 4
Describing the Data Set..... 7

2.1 Selecting a Data Set

☑ To Select a Data Set

1. In the list *Data Type*, select the data set file type.
2. Use the *Browse* button located on the right of the *Folder* field to select the folder or database containing the data set.
3. Use the *Browse* button located on the right of the *Data Set* field to select the data set file or table.



4. Click the *Next* button.

   If the **data source is stored in a Oracle, Teradata or SQL Server 2005 data base**, an additional warning may be displayed, asking you to validate the execution of the query. For more information, refer to chapter Special Case: Data Stored in Databases (on page 5).

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

2.1.1 Special Case: Data Stored in Databases - the Explain Mode

Before requesting data stored in a Teradata(1), Oracle(2) or SQLServer 2005 database, SAP InfiniteInsight® uses a feature, called the **Explain mode**, which categorizes the performances of SQL queries in several classes defined by the user. In order to be as fast and as light as possible, this categorization is done **without actually executing the full SQL query**.

> **Note**
>
> (1) For all versions of Teradata.
> (2) For all versions above and including Oracle 10.

The objective is to allow you to estimate the workload of the SQL query before executing it and then deciding --possibly thanks to an IT Corporate Policy-- if the SQL query can actually be used.

For example, an IT Corporate Policy may favor interactivity and then define 3 classes of SQL queries, each with its maximum time:

- *Immediate*: *duration < 1 s*. The query is accepted and executed immediately.
- *Batched*: *1s <= duration < 2 s*. The query is accepted but will be executed on next idle time.
- *Rejected*: *2s <= duration*. The query will never be executed.

The number, names and limits of classes are defined by the user in order for these values to match the current DBMS configuration and DBMS usage policy.

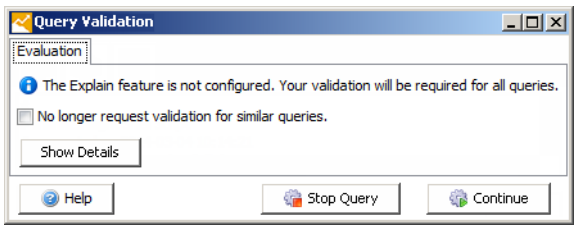
The Explain Mode has been ConfiguredIf the Explain mode has been configured by your DBMS administrator, there are two possible outcomes to a query:

- **the query is accepted and executed**: this is completely transparent. SAP InfiniteInsight® accesses the data without further input from the user.
- **the query needs to be validated before being executed**: a pop-up window opens displaying a message configured by the DBMS administrator. A query that needs validation can be categorized in two ways:
  - **medium-sized**
    You will probably have to check with your administrator to know which action to take:
- If the administrator authorizes the query, click the *Continue* button. The pop-up window closes and the requested action is carried out.
- If the administrator does not authorize the query, click the button *Stop Query*, the pop-up window closes, but no action is executed.
  - **huge**
    It means that the query will take too much time and resources. In that case, the behavior of the *Continue* button depends on the configuration set by the DBMS Administrator (for example, it can automatically refuse queries that are considered too heavy). In any case, you should check with them to know the line of action to follow.

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

The Explain Mode has not been Configured [#]

If your DBMS Administrator has not configured the Explain mode, the following pop-up opens when you try to access the data:



You need to contact your Administrator who will tell you which action to take and configure the Explain mode.

If the Administrator validates the execution of the query, you may want all queries with the same duration to be executed without validation. In that case, check the box *Do not request validation anymore for similar requests*. The validation message will then only appear for larger queries. This configuration will only be used for the current session, when closing SAP InfiniteInsight®, it will be lost. For a permanent configuration, see your DBMS Administrator, who will find the necessary information in the support document Explain Mode available in section Support and Integration Documentation of SAP InfiniteInsight® documentation.

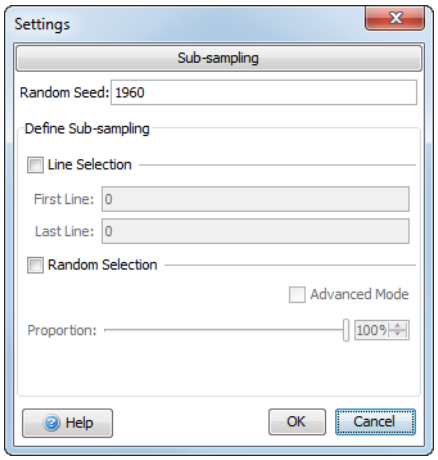
2.1.2 Sub-sampling a Data Set

Sub-sampling can be achieved by:

- selecting a specific first line and last line from the original data,
- randomly selecting a proportion of the original data. A random seed can be used to select the same records from the original data when sub-sampling,
- mixing the selection by line and the random selection.

☑ To Define the Sub-sampling Settings

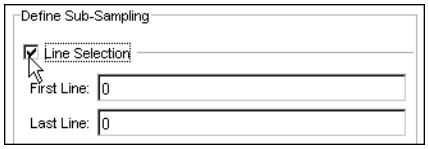
Click the button (*Advanced Settings*) located on the right of the field *Data Set*. A pop-up window opens, allowing you to set the sub-sampling parameters.



© 2014 SAP SE or an SAP affiliate company. All rights reserved-

☑ To Set a Line Selection

1 Check the option *Line Selection* in the section *Define Sub-sampling*.

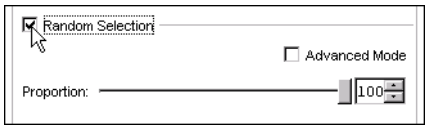


2 In the field *First Line*, enter the number of the line you want your sub-sampled data set to begin with.

3 In the field *Last Line*, enter the number of the line you want your sub-sampled data set to end with.

☑ To Set a Random Sub-sampling

4 Check the option *Random Selection* in the section *Define Sub-sampling*.



5 In the section *Random Selection*, use the *Proportion* slide to select the percentage of the original data set you want to keep.

6 In the field *Random Seed*, enter a seed value for the sub-sampling. This step is optional.

7 Use the *Proportion* slide to select the percentage of the original data set you want to keep.

2.2 Describing the Data Set

To describe your data, you can:

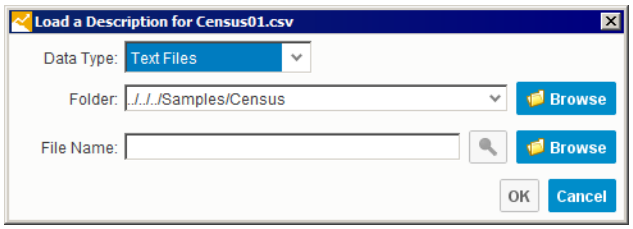
- Either **use an existing description file**, that is, taken from your information system or saved from a previous use of SAP InfiniteInsight® features,
- Or **create a description file using the** *Analyze* **option**, available to you in InfiniteInsight® modeling assistant. In this case, **it is important that you validate the description file obtained**. You can save this file for later re-use. If you name the description file **KxDoc_<SourceFileName>**, it will be automatically loaded when clicking the *Analyze* button.

Important - The description file obtained using the *Analyze* option results from the analysis of the first 100 lines of the initial data file. In order to avoid all bias, we encourage you to mix up your data set before performing this analysis.

For more information on data set description, see **InfiniteInsight® - User Guide**.

☑ To Load a Description File

1 On the screen *Data Description*, click the button *Open Description*. The following window opens:



© 2014 SAP SE or an SAP affiliate company. All rights reserved-

2 In the window *Load a Description*, select the type of your description file.

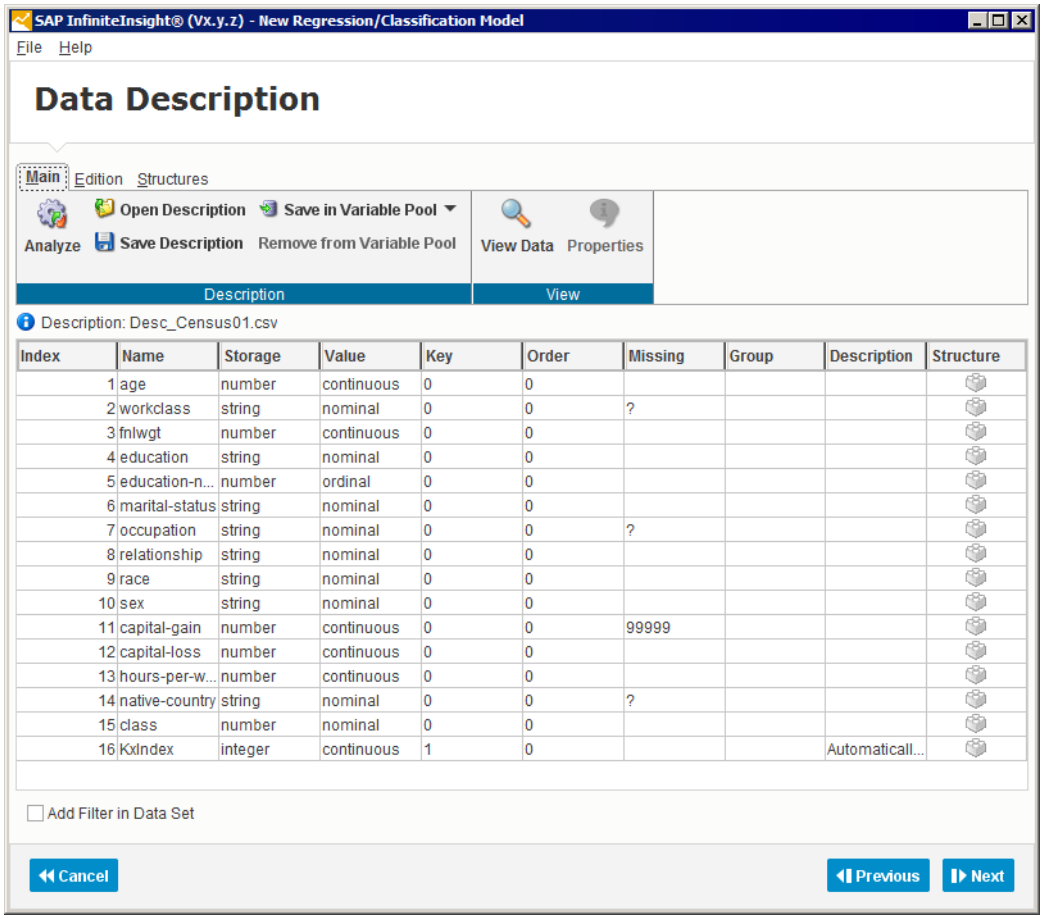
3 In the *Folder* field, select the folder where the description file is located with the *Browse* button.

> **Note**
>
> The folder selected by default is the same as the one you selected on the screen *Data to be Modeled*.

1 In the *Description* field, select the file containing the data set description with the *Browse* button.

> **Caution**
>
> When the space used for model training **contains** a physical variable named `KxIndex`, it is not possible to use a description file without any key for the described space.
> When the space used for model training **does not contain** a physical variable named `KxIndex`, it is not possible to use a description file including a description about a KxIndex variable since it does not exist in current space.

1 Click the *OK* button. The window *Load a Description* closes and the description is displayed on the screen *Data Description*.



Index	Name	Storage	Value	Key	Order	Missing	Group	Description	Structure
1	age	number	continuous	0	0				
2	workclass	string	nominal	0	0	?			
3	fnlwgt	number	continuous	0	0				
4	education	string	nominal	0	0				
5	education-n...	number	ordinal	0	0				
6	marital-status	string	nominal	0	0				
7	occupation	string	nominal	0	0	?			
8	relationship	string	nominal	0	0				
9	race	string	nominal	0	0				
10	sex	string	nominal	0	0				
11	capital-gain	number	continuous	0	0	99999			
12	capital-loss	number	continuous	0	0				
13	hours-per-w...	number	continuous	0	0				
14	native-country	string	nominal	0	0	?			
15	class	number	nominal	0	0				
16	KxIndex	integer	continuous	1	0			Automaticall...	

2 Click the *Next* button.

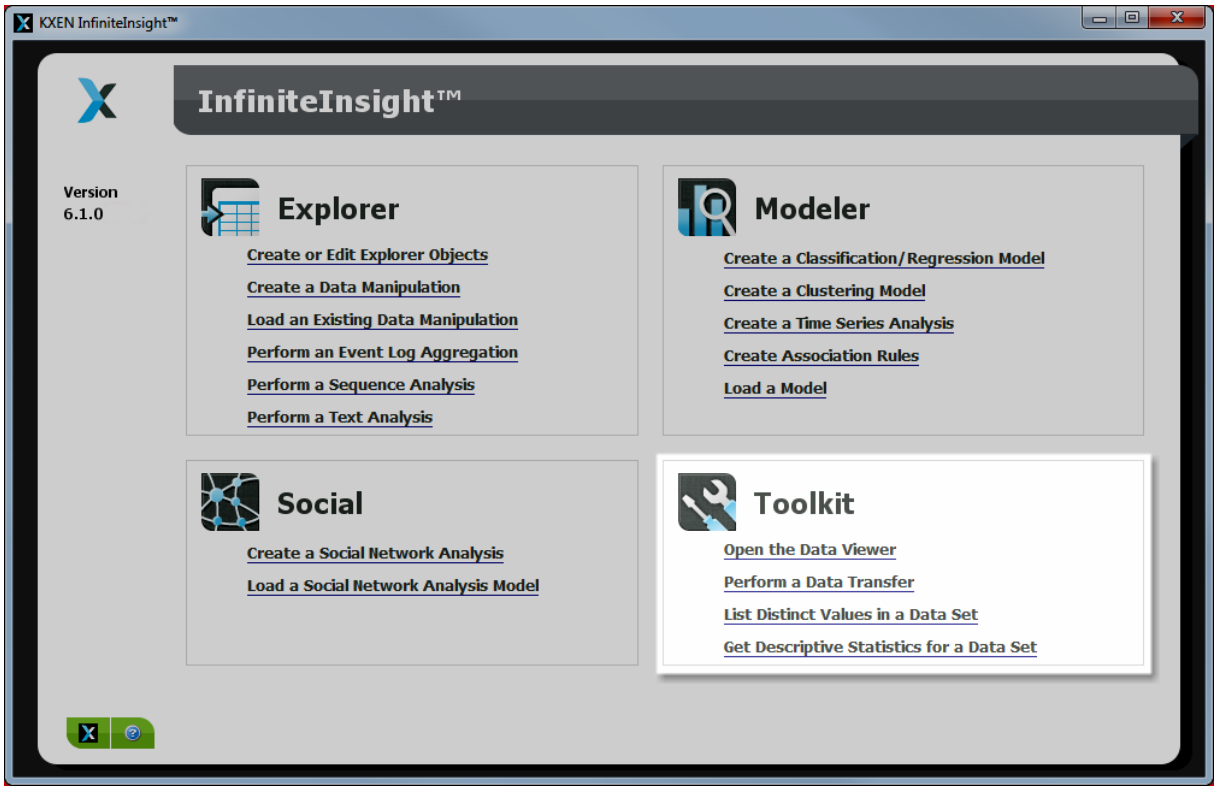
© 2014 SAP SE or an SAP affiliate company. All rights reserved-

3 Data Viewer

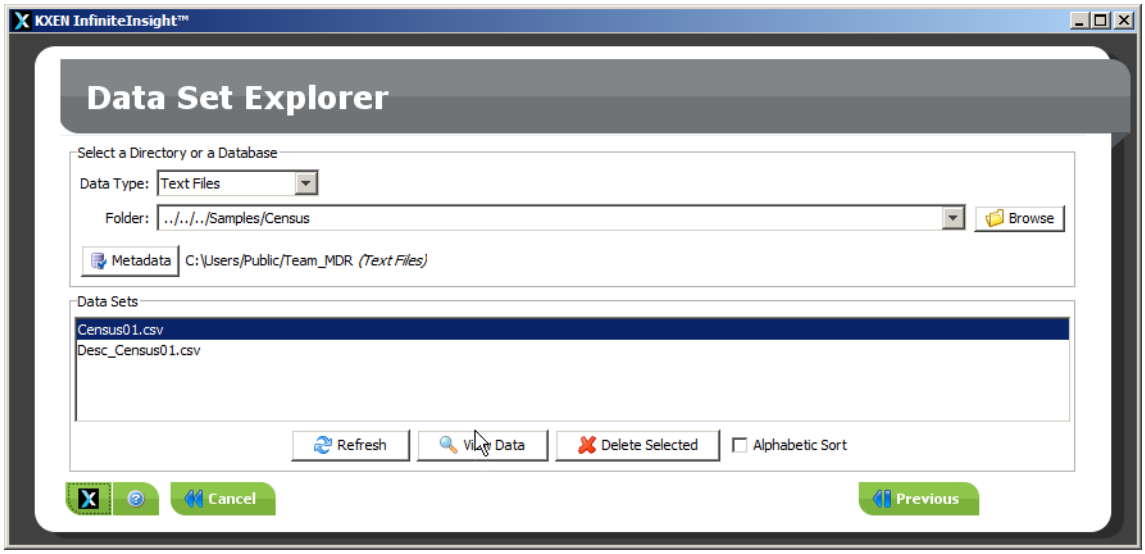
The *Data Viewer* allows you to visualize existing data sets and if needed to delete them.

☑ To Display the Data Sets List

1 On the *Start* panel, click the feature *Open the Data Viewer* in the section *Toolkit*.



The panel allowing you to select the location where your data sets are stored is displayed.



© 2014 SAP SE or an SAP affiliate company. All rights reserved-

2 In the list *Data Type*, select the type of storage the data sets you want to browse are located in (Text File, Database, ...).

3 Use the *Browse* button located on the left of the *Folder* field to select the folder (or database) containing the data sets. If you want to explore data sets stored in a password protected database, do not forget to provide the user name and password granting you access to the database.

The data sets list is displayed in the field *Data Sets*.

IN THIS CHAPTER

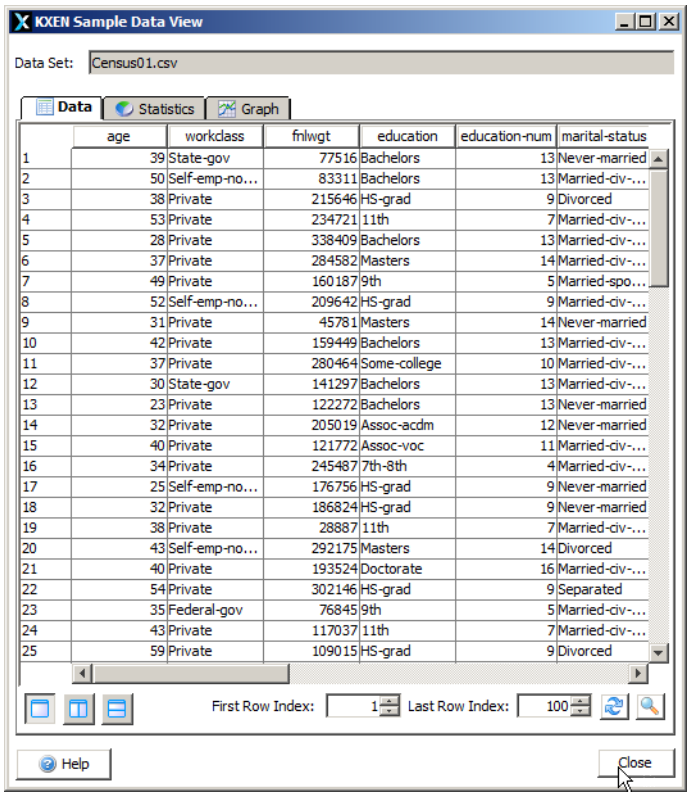
Visualizing a Data Set 10
Deleting a Data Set 11

3.1 Visualizing a Data Set

☑ To Visualize a Data Set

1 Click the data set you want to visualize. The button *View Data* at the bottom of the panel becomes active.

2 Click the button *View Data*. The first hundred lines in the data set are displayed in a new window.



	age	workclass	fnlwgt	education	education-num	marital-status
1	39	State-gov	77516	Bachelors	13	Never-married
2	50	Self-emp-no...	83311	Bachelors	13	Married-civ-...
3	38	Private	215646	HS-grad	9	Divorced
4	53	Private	234721	11th	7	Married-civ-...
5	28	Private	338409	Bachelors	13	Married-civ-...
6	37	Private	284582	Masters	14	Married-civ-...
7	49	Private	160187	9th	5	Married-spo...
8	52	Self-emp-no...	209642	HS-grad	9	Married-civ-...
9	31	Private	45781	Masters	14	Never-married
10	42	Private	159449	Bachelors	13	Married-civ-...
11	37	Private	280464	Some-college	10	Married-civ-...
12	30	State-gov	141297	Bachelors	13	Married-civ-...
13	23	Private	122272	Bachelors	13	Never-married
14	32	Private	205019	Assoc-acdm	12	Never-married
15	40	Private	121772	Assoc-voc	11	Married-civ-...
16	34	Private	245487	7th-8th	4	Married-civ-...
17	25	Self-emp-no...	176756	HS-grad	9	Never-married
18	32	Private	186824	HS-grad	9	Never-married
19	38	Private	28887	11th	7	Married-civ-...
20	43	Self-emp-no...	292175	Masters	14	Divorced
21	40	Private	193524	Doctorate	16	Married-civ-...
22	54	Private	302146	HS-grad	9	Separated
23	35	Federal-gov	76845	9th	5	Married-civ-...
24	43	Private	117037	11th	7	Married-civ-...
25	59	Private	109015	HS-grad	9	Divorced

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

☑ To Select the Number of Lines to Display

3 In the field *First Row Index*, enter the index of the line before the first one you want to display (for example, to start on line 1, enter 0).

4 In the field *Last Row Index*, enter the index of the line before the last one you want to display (for example, to end on line 100, enter 99).

5 Click the *Refresh* button to display the selected lines.

☑ To Search for a Variable

6 Click the *Search* button (Magnifier icon). The window *Search Variable by* opens.

7 Select the type of search you want to perform (by index or by name). A variable index is its column number in the data set.

8 In the displayed drop-down list, select the variable index or name.

9 Click the *OK* button. The searched variable column appears highlighted in the window *KXEN Sample Data View*.

☑ To Display the Variables Statistics for the First Data Set Lines

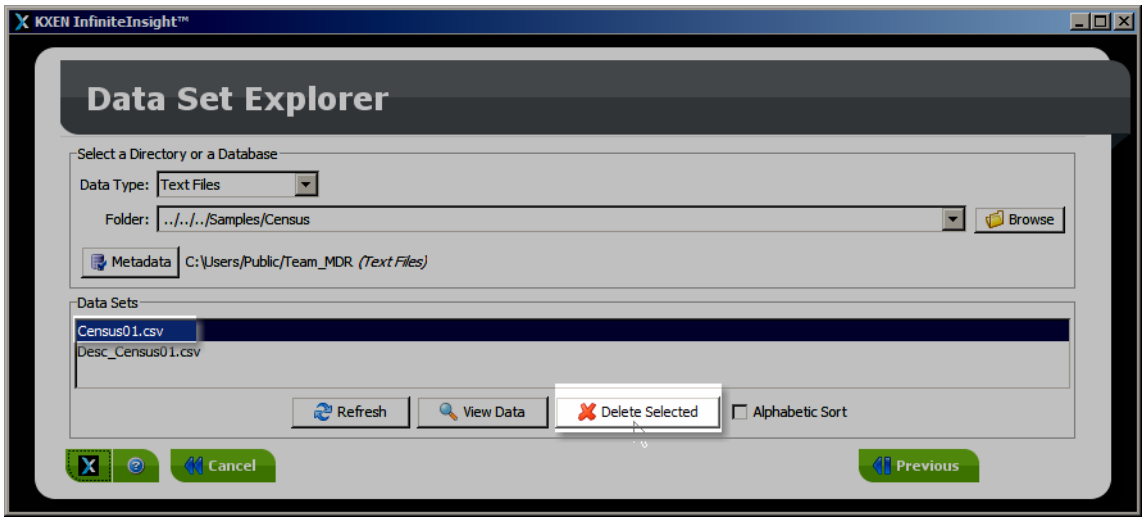
10 In the field *Number of Lines*, select the number of lines for which you want to see the statistics. To display the statistics for the whole data set, enter a number below 1.

11 Click the *Statistics* button (graph icon). The window *Statistics for "<DataSetName>" on the first <n> Lines* opens. For more details on statistics, refer to section Descriptive Statisctics > From the Panel View Data (see "From the Panel "View Data"" on page 29).

3.2 Deleting a Data Set

☑ To Delete a Data Set

1 Select the data set you want to delete. The button *Delete Selected* becomes active.



2 Click the button *Delete Selected*. A dialog box is displayed asking you to confirm the deletion.

3 Click the *Yes* button. The data set is deleted.

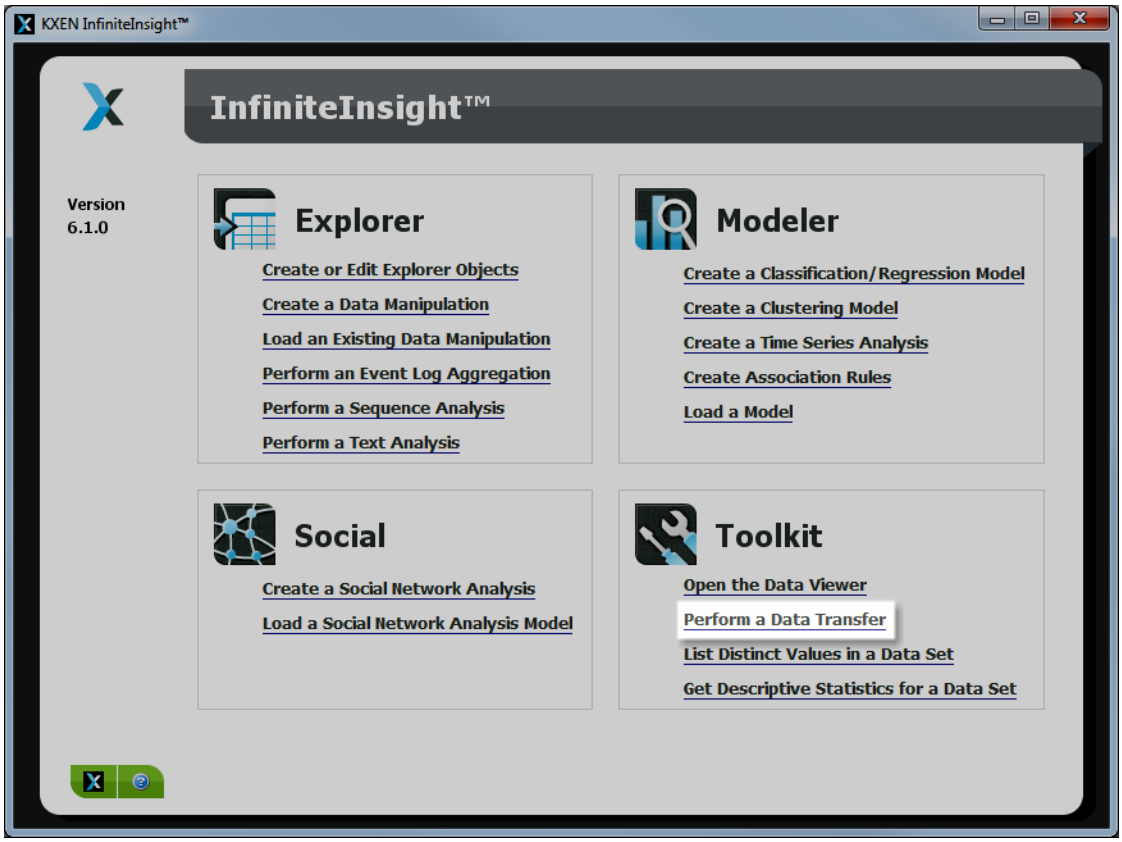
© 2014 SAP SE or an SAP affiliate company. All rights reserved-

4 Data Transfer

The component *Data Transfer* allows you to transfer a data source to another location or format.

☑ To Start a Data Transfer

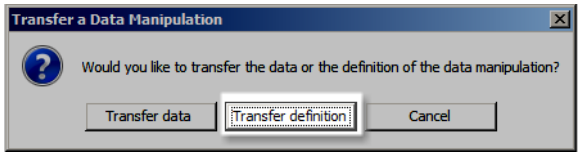
1 On the *Start* panel, click the option *Perform a Data Transfer* in the *Toolkit* section.



The panel allowing you to select the data set you want to transfer is displayed.

2 Select the data set you want to transfer (see Selecting a Data Set (see "Selecting a Data Set [#T]" on page 4)).

If you want to **transfer an analytical data set,** a message box will be displayed after clicking the *Next* button, asking you whether you want to transfer the data contained in the analytical data set or if you want to transfer only its definition.


Transfer a Data Manipulation
Would you like to transfer the data or the definition of the data manipulation?
Transfer data
Transfer definition
Cancel

- If you choose to transfer the data, the data transfer component will act as if it was a classical data set, and ask you to describe the data (see "Descriptive Statistics" on page 20) and to save the new data set (see "Saving the New Data Set" on page 15).
- If you choose to transfer only the analytical data set definition, the panel *Data Manipulation Transfer* will be displayed (see Transferring an Analytical Data Set (see "Transferring a Data Manipulation" on page 13)).

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

IN THIS CHAPTER

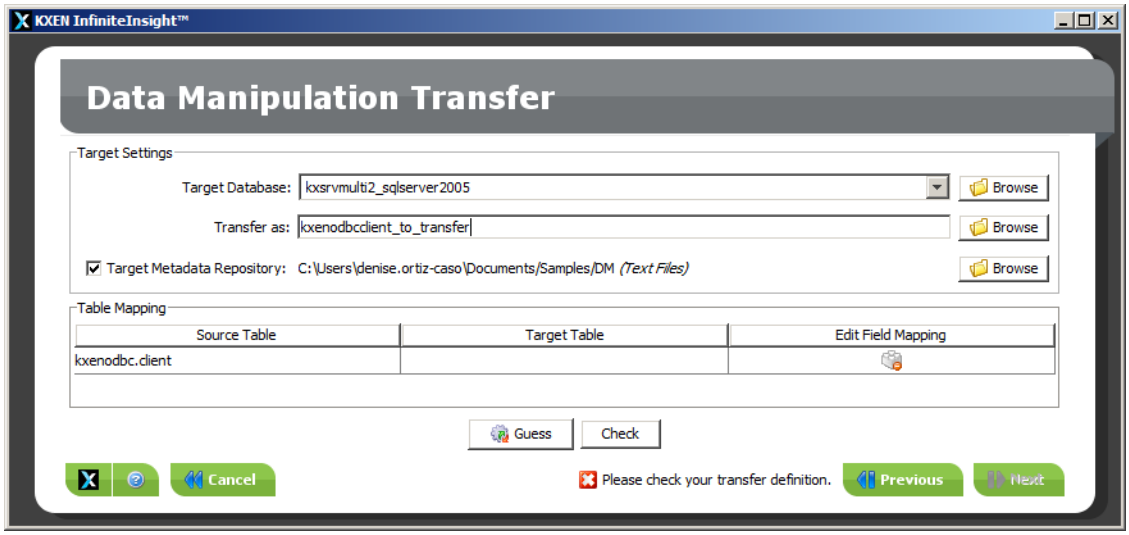
Transferring a Data Manipulation 13
Selecting Variables 14
Saving the New Data Set 15

4.1 Transferring a Data Manipulation

When transferring a data manipulation, the panel *Data Manipulation Transfer* allows you to select the target database, to define the mapping between the source tables and the target tables and, if needed, to define the mapping between the table fields.

☑ To Define a Transfer

1. To select the target database (that is, the database in which you want to transfer the definition of the analytical data set), click the *Browse* button, then select a database in the displayed dialog box. Click the *OK* button to validate your selection.
2. In the field *Transfer as*, enter the name you want the analytical data set to be identified by.



3. Click the *Guess* button, located on the lower part of the panel. InfiniteInsight® modeling assistant tries to find in the target data base, the tables used in the analytical data set definition based on their name. For more information on field mapping, see section Mapping Fields. (see "Mapping Fields" on page 14)
4. For each table not found or incorrectly identified, click the corresponding cell in the column *Target Table*. A table selector is displayed.
5. When all the tables are mapped, click the *Check* button. InfiniteInsight® modeling assistant checks if the mapping is correct for all tables. If an error is found, a diagnostic message is displayed, else the *Next* button is enabled.
6. Click the *Next* button to start the transfer.

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

4.1.1 Mapping Fields

When the fields from the target table do not have the same names as the fields from the source table, you need to map the source fields to the target fields.

The column *Edit Field Mapping* indicates the state of the field mapping.

- : the transfer definition has not yet been validated or the default mapping is correct.
- : the mapping has been edited and is correct.
- : the mapping is incorrect (whether it has been edited or not).

☑ To Define the Field Mapping

1. In the column *Edit Field Mapping*, click the icon . A window opens allowing you to edit the field mapping.
2. Click the Guess button. InfiniteInsight® modeling assistant tries to map the fields based on a name similarity and on the fields storage format.

   Note - Two fields can be mapped together only if they have the same storage format.
3. To modify a mapping that seems incorrect or that has not been found:
   1. On the line corresponding to the mapping to modify, click the *Target Field* column.
   2. In the drop-down list, select the target field you want to map with the source field.
4. When all the fields are mapped, click the *Check* button. If all the fields are correctly mapped, the *OK* button is enabled, else a diagnostic message is displayed.
5. Click the *OK* button to validate the mapping.

4.2 Selecting Variables

The panel Variable Selection allows you to choose which variables you want to exclude from the data set.

☑ To Exclude a Variable

1. Select the variable to exclude in the Variables Selected for Transfer list.
2. Click on the > button located on the left of the Excluded Variables screen section.

Note - All the lists can be sorted by the order in the file or alphabetically.

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

4.3 Saving the New Data Set

This panel allows you to duplicate a data set into a file or a data base table.

☑ To Save the Data Set

1. Use the *Browse* button located on the right of the *Folder* field to select the folder or data base where the file will be saved.
2. Enter the name of the file to create or use the *Browse* button located on the right of the field *Output Dataset* to select an existing file.
3. Click the *Next* button.

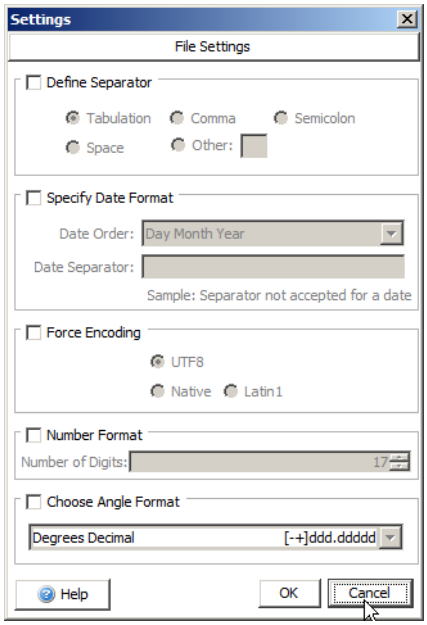
   A message box appears displaying a progress bar. When the file has been created, the main menu is displayed.

4.3.1 Defining Text File Settings

When transferring the data set to a text file, you will be able to define various specific settings such as the column separator, the date format, the number format or the encoding.

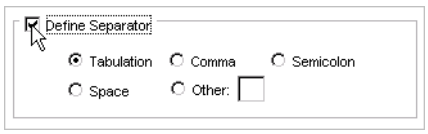
☑ To Define the File Settings

1. Click the button (*Advanced Settings*). A pop-up window opens allowing you to define the text file settings.



☑ To Define the Separator

2. Check the box *Define Separator*. The corresponding section is activated.

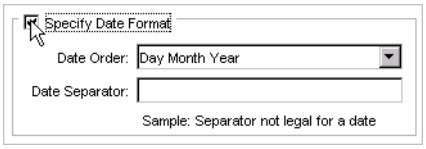


© 2014 SAP SE or an SAP affiliate company. All rights reserved-

3 Select the character to be used as the column separator in the generated text file. There are four predefined choices (*Tabulation*, *Comma*, *Semicolon*, and *Space*), but you can define your own separator by selecting the *Other* option and entering the character of your choice in the provided field.

☑ To Specify the Date Format

4 Check the box *Specify Date Format*. The corresponding section is activated.

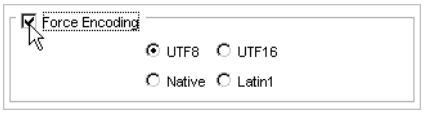


5 In the list *Date Order*, select in which order you want the day, month and year to be displayed.

6 In the field *Date Separator*, enter the character(s) that will appear between the different date parts. An example is displayed below the field *Date Separator*.

☑ To Force the Encoding

7 Check the box *Force Encoding*. The corresponding section is activated.

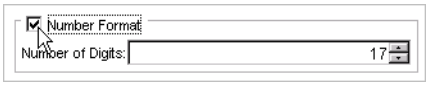


8 Select the encoding you want to use for the generated text file. Four types of encodings are available:

- *UTF8*
- *UTF16*
- *Native*
- *Latin1*

☑ To Define the Number Format

9 Check the box *Number Format*. The corresponding section is activated.



10 In the field *Number of Digits*, enter the number of digits to display after the decimal.

☑ To Define an Angle Format

11 Check the box *Choose Angle Format*. The corresponding section is activated.

12 Select one of the four avaible options:

- Degrees Decimal
- Degrees Minutes Decimal
- Degrees 1000 Minutes Decimal (NMEA)
- Degrees Minutes Seconds
-

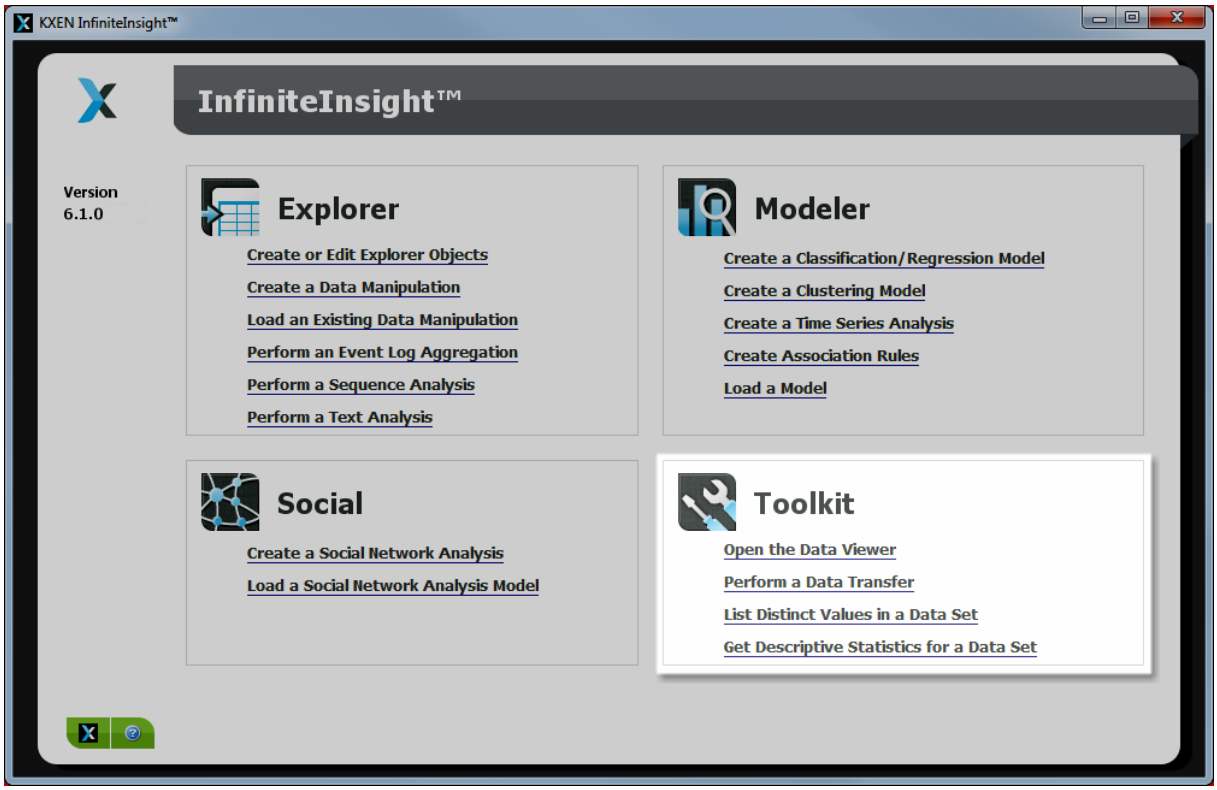
© 2014 SAP SE or an SAP affiliate company. All rights reserved-

5 List Distinct Values

The feature *List Distinct Values in a Data Set* allows you to export the list of distinct values of a variable from your data set. Among other things, it allows you to generate a list of unique index values from an existing data set and to use it as a reference table in InfiniteInsight® Explorer - Event Logging or InfiniteInsight® Explorer - Sequence Coding.

☑ To Use the Feature List Distinct Values

1 On the *Start* panel, click the option *List Distinct Valuesin a Data Set* in the section *Toolkit*.



The panel allowing you to select the data set from which the values will be extracted is displayed.

2 For a description of how to select and describe the data set, refer to the section Common Steps (on page 4), then go on with reading the sections below.

IN THIS CHAPTER

Selecting the Variable to Extract 18
Saving the Categories 18

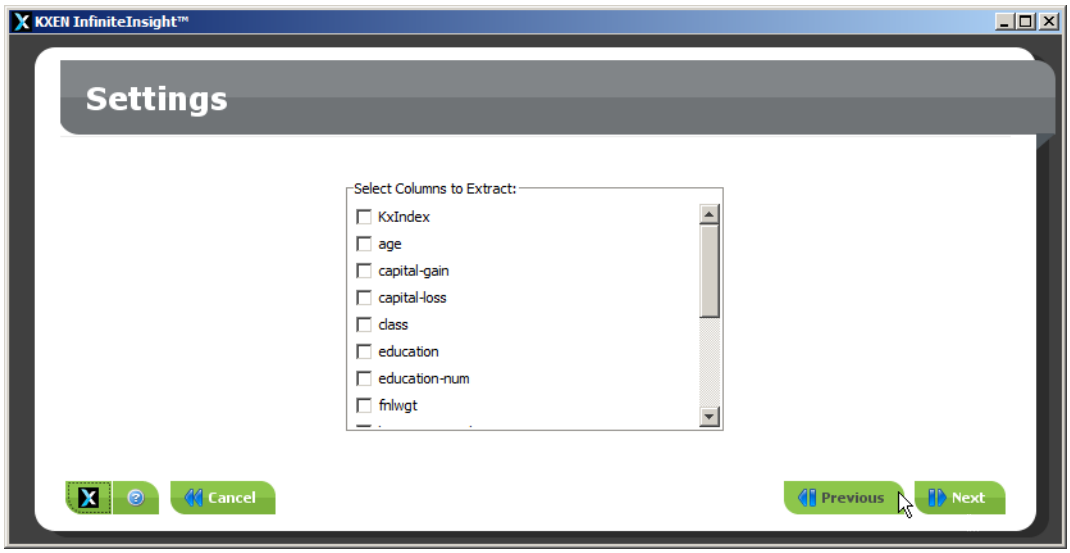
© 2014 SAP SE or an SAP affiliate company. All rights reserved-

5.1 Selecting the Variable to Extract

The *Settings* panel allows you to select the variable for which you want to extract the categories.

☑ To Select a Variable

1 Check the variable in the list *Select Columns to Extract*.



2 Click the *Next* button.

5.2 Saving the Categories

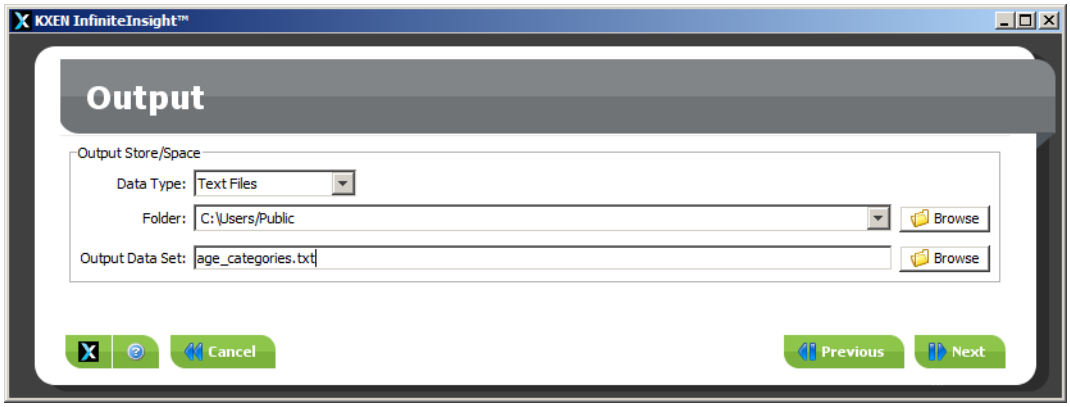
1 The next panel allows you to save the variable categories extracted from the data set.

☑ To Save the Variable Category List

2 In the list *Data Type*, select in which format you want to save the category list.

3 Use the *Browse* button located on the right of the *Folder* field to select the folder or data base where the file will be saved.

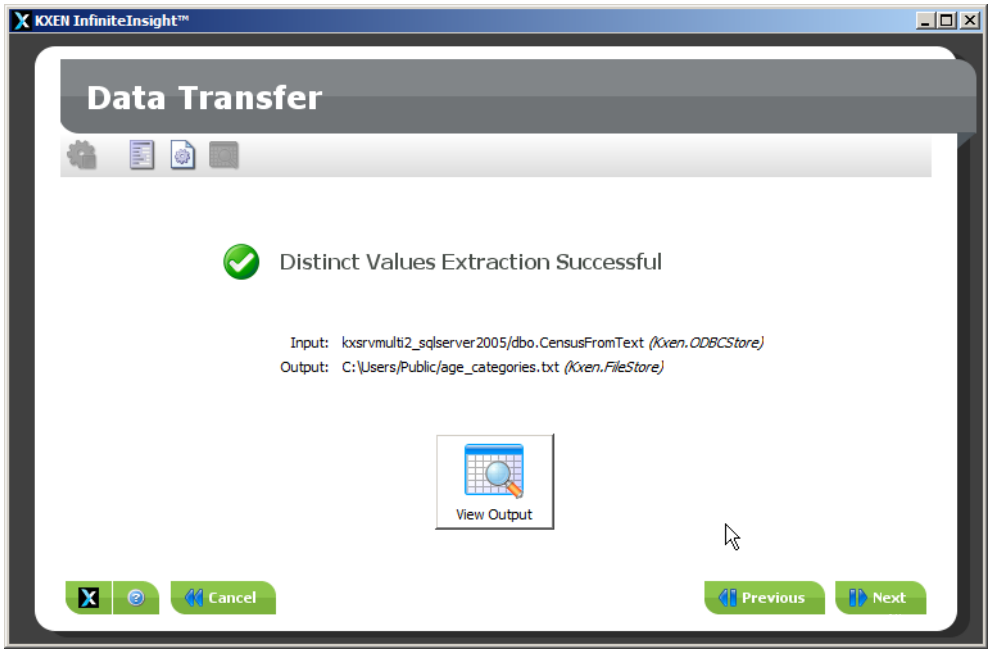
4 Enter the name of the file or table to create or use the *Browse* button located on the right of the field *Output Data Set* to select an existing file or table.



Click the *Next* button. A progress bar is displayed allowing you to follow the process.

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

5 When the file has been created successfully, the following panel is displayed.
Click *View Output*.



6 Click the *Next* button to go back to the *Start* panel.

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

6 Descriptive Statistics

This component allows you to generate variables statistics for a data set, such as the cross-statistics or the profit curves when a variable in the data set is used to estimate the target variable. It can be accessed either through the option *Descriptive Statistics* of the *Start* panel or by using the option *View Data* provided when selecting a data source in any KXEN component.

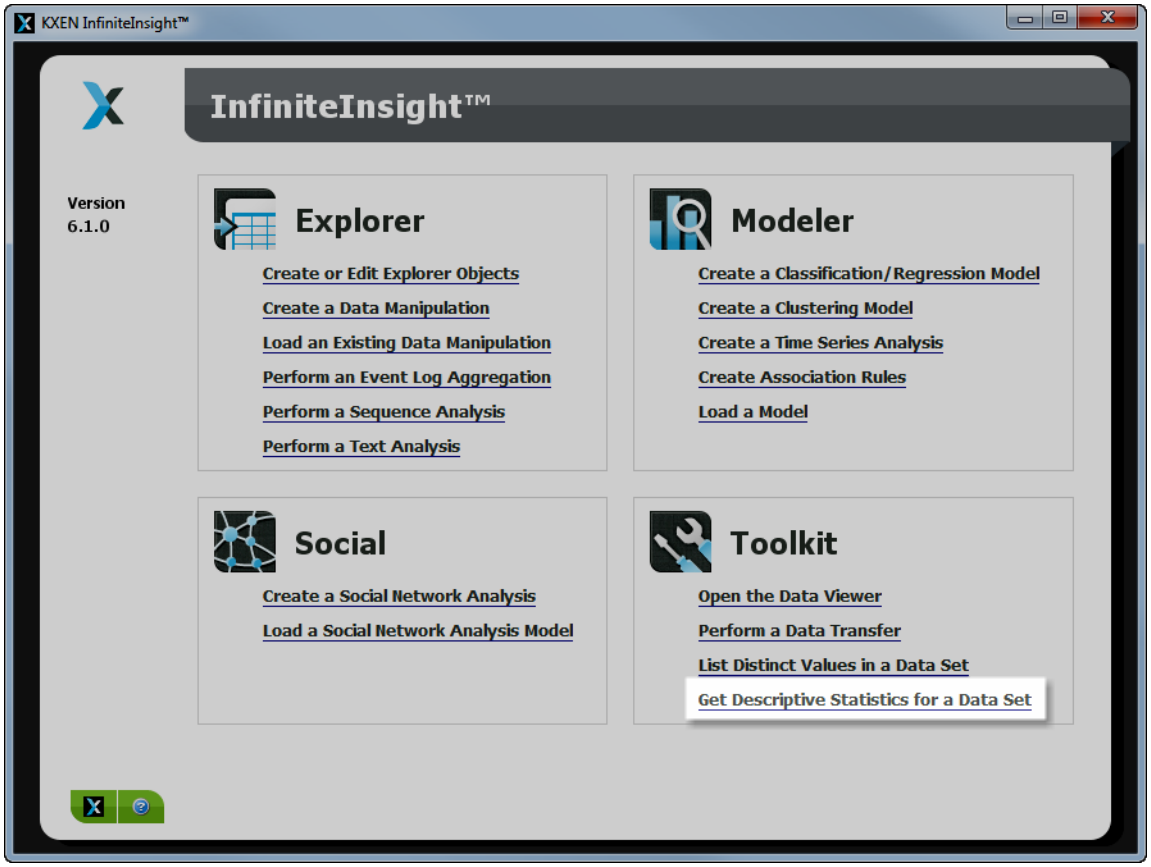
IN THIS CHAPTER

From the Start Panel 20
From the Panel "View Data" 29
Geo-location 32

6.1 From the Start Panel

☑ To Access the Descriptive Statistics from the Start Panel

1 On the *Start* panel, click the option *Get Descriptive Statistics for a Data Set* in the *Toolkit* section.



© 2014 SAP SE or an SAP affiliate company. All rights reserved-

The panel allowing you to select the data set you want to analyze is displayed.

2 For a description of how to select and describe a data set, refer to the section Common Steps (on page 4), then go on with reading the sections below.

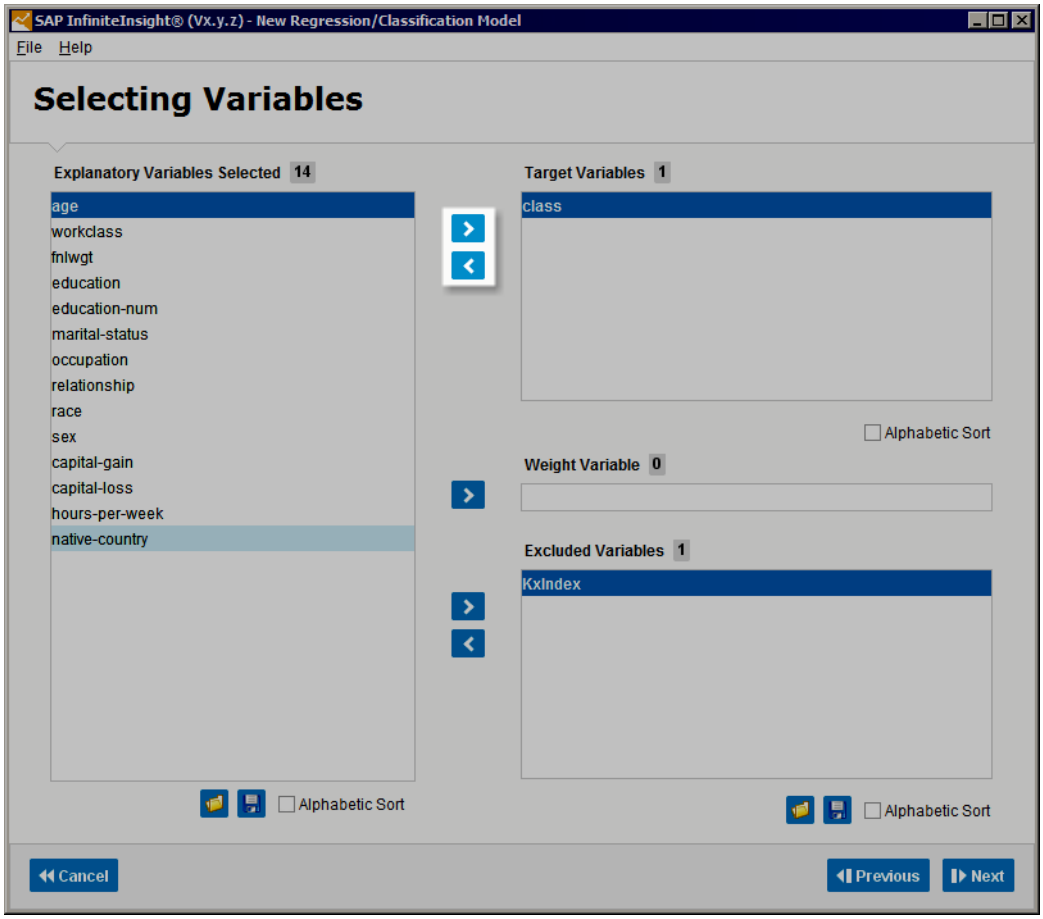
6.1.1 Selecting the Variables

Once the training data set and its description have been entered, you can select different variables :

- one or more Targets Variables,
- possibly a Weight Variable,
- and the Explanatory Variables.

☑ To Select Targets Variables

3 On the screen *Selecting Variables*, in the section *Explanatory variables selected* (left hand side), select the variables you want to use as Target Variables.



Note

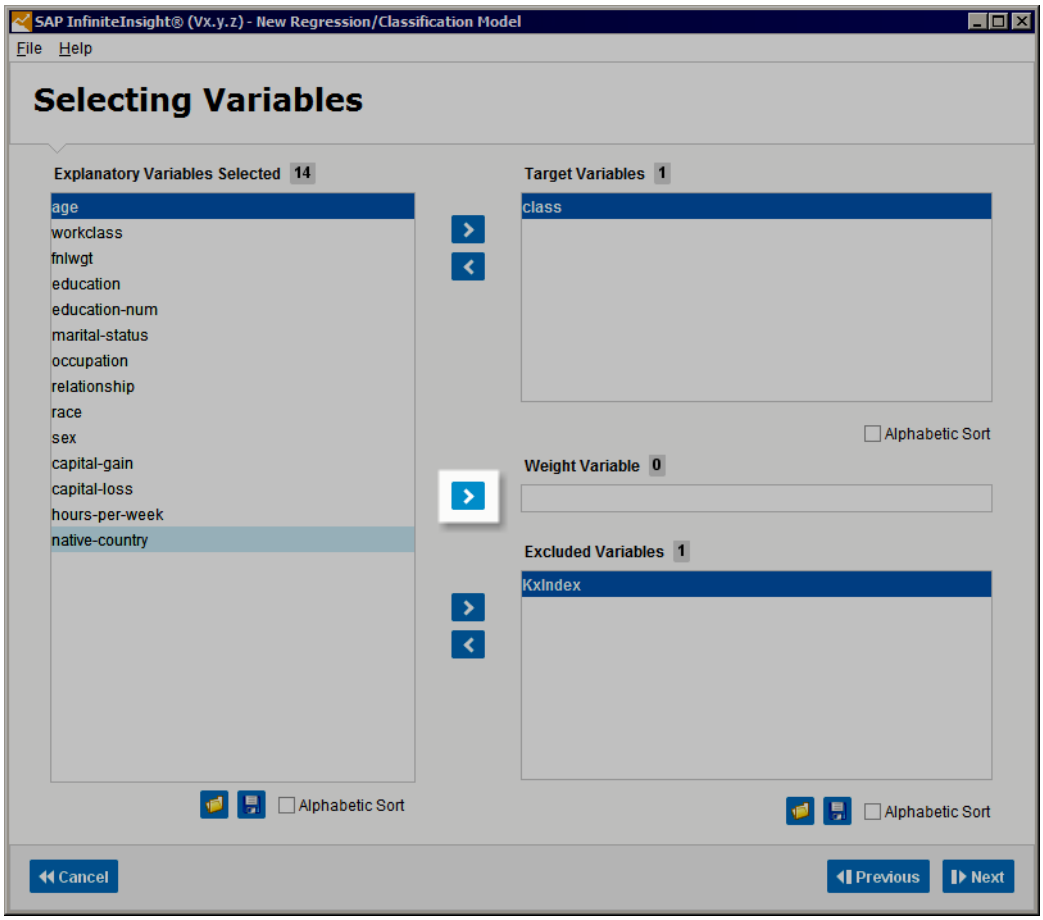
On the screen *Selecting Variables*, variables are presented in the same order as that in which they appear in the table of data. To sort them alphabetically, select the option *Alphabetic sort*, presented beneath each of the variables list.

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

4 Click the button > located on the left of the screen section *Target Variables* (upper right hand side).
The variable moves to the screen section *Target Variables*.
Also, select a variable in the screen section *Target Variables* and click the button < to move the variables back to the screen section *Explanatory variables selected*.

☑ To Select a Weight Variable

1 On the screen *Selecting Variables*, in the section *Explanatory variables selected* (left hand side), select the variables you want to use as a Weight Variable.



Note

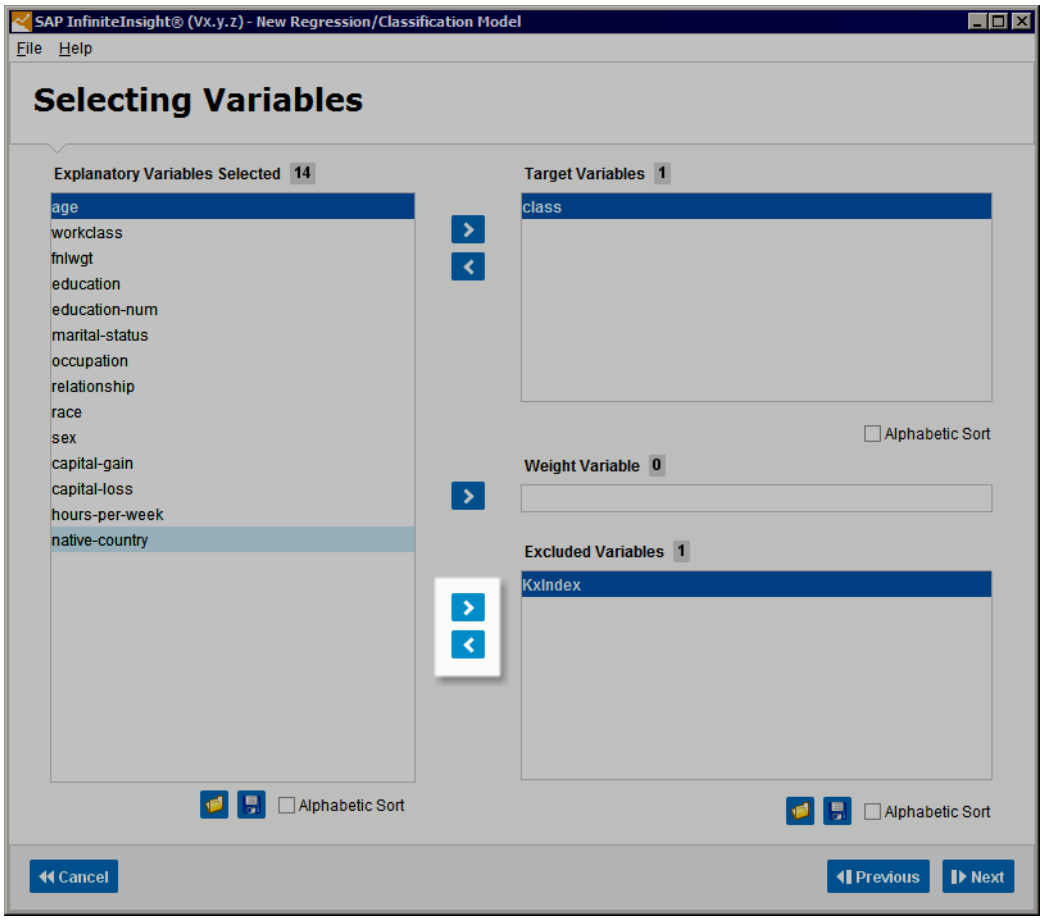
On the screen *Selecting Variables*, variables are presented in the same order as that in which they appear in the table of data. To sort them alphabetically, select the option *Alphabetic sort*, presented beneath each of the variables list.

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

2 Click the button > located on the left of the screen section *Weight Variable* (middle right hand side).
The variable moves to the screen section *Weight Variable*.
Also, select a variable in the screen section *Weight Variable* and click the button < to move the variables back to the screen section *Explanatory variables selected*.

☑ To Exclude some Variables from Data Analysis

1 On the screen *Selecting Variables*, in the section *Explanatory Variables Selected* (left hand side), select the variable to be excluded.



Note

On the screen *Selecting Variables*, variables are presented in the same order as that in which they appear in the table of data. To sort them alphabetically, select the option *Alphabetic sort*, presented beneath each of the variables list.

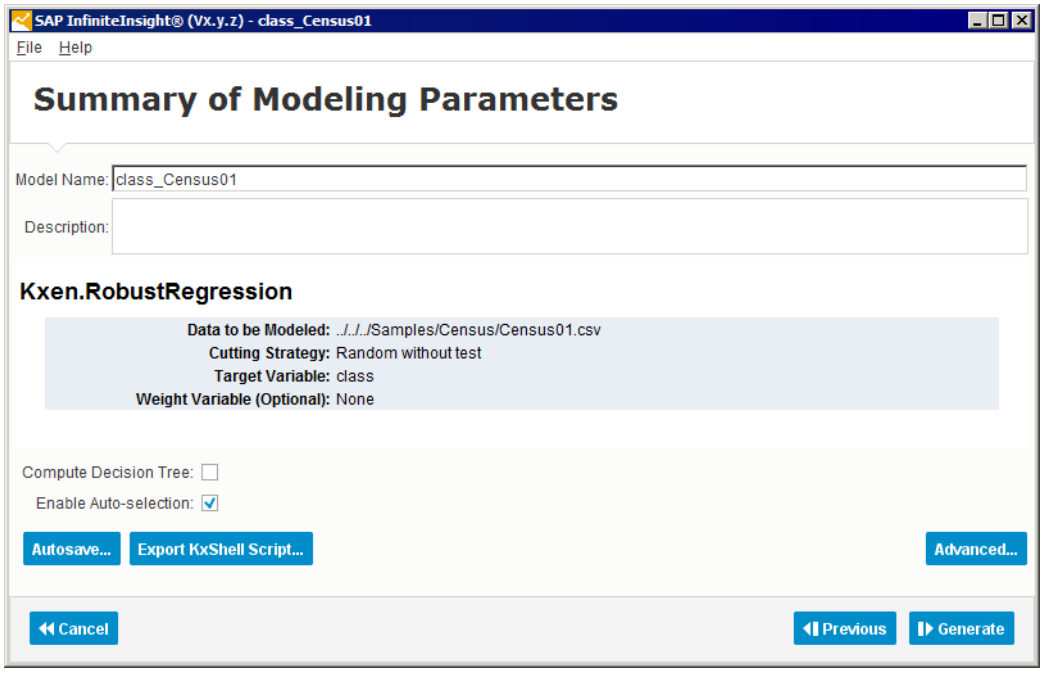
1 Click the button > located on the left of the screen section *Variables excluded* (lower right hand side).
The variable moves to the screen section *Variables excluded*.
Also, select a variable in the screen section *Variables excluded* and click the button < to move the variables back to the screen section *Explanatory variables selected*.

2 Click the *Next* button.

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

6.1.2 Checking the Parameters

The screen *Summary of Modeling Parameters* allows you to check the modeling parameters just before generating the model.



Options	Description
Model Name	The **name** of model is filled automatically. It corresponds to the name of the target variable, followed by the underscore sig ("_") and the name of the data source, minus its file extension.
Description	Once the model is saved, the **description** will allow you to recognize your model when you want to reopen it with SAP InfiniteInsight®.
Autosave...	This button opens the panel *Model Autosave* allowing you to indicate that the model must be automatically saved at th end of the learning process and ot set all the required parameters. When the autosave option is activated, a green check ma is displayed on the *Autosave* button. At then end of the learning process, a warning message will be displayed confirmi that the model has been saved
Export KxShell Script...	This button opens the panel allowing you to generate a KxShell script reproducing the current model.
Advanced...	This button allows you to set estimator variables for the target(s) and/or the clusters. For more details, see section **Advanced Parameters** (on page 25).

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

Advanced Parameters

Target Estimator Variables

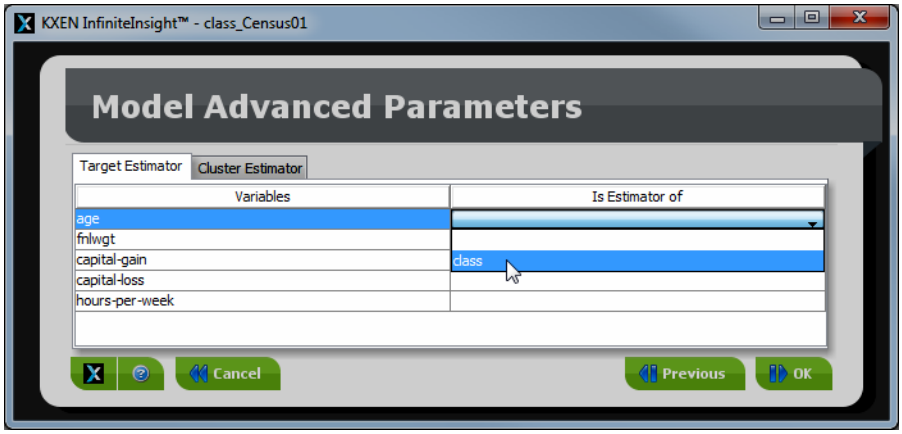
If you have selected at least one target variable, the tab *Define Estimator* will be activated, allowing you to select variables as estimator of the target variables. An **estimator** is a variable that can be used to **predict the target** variable behavior. Using an estimator is equivalent to creating a model.

If there is no variable that can be used as estimator of the target in your data set, do not fill anything in this panel.

Note - Only continuous variables can be used as estimators.

☑ To Select an Estimator of the Target

1. Select the tab *Target Estimator*.
2. In the *Variables* column, click the variable you want to use as an estimator.
3. Click the corresponding cell in the column is *Estimator of*. A list of available targets is displayed.



4. Select the target to be estimated.
5. Repeat steps 2 to 4 for all the estimator variables.
6. Click the *Next* button.

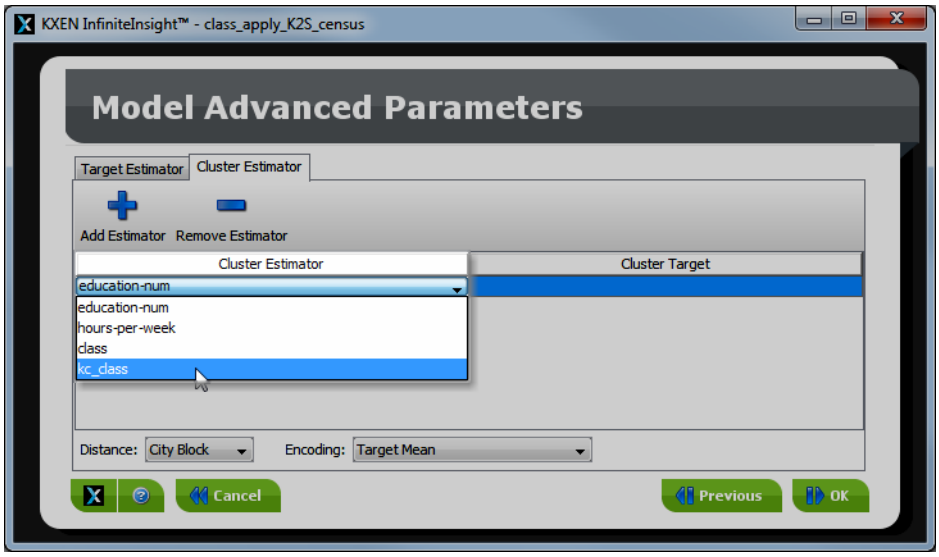
© 2014 SAP SE or an SAP affiliate company. All rights reserved-

Cluster Estimator

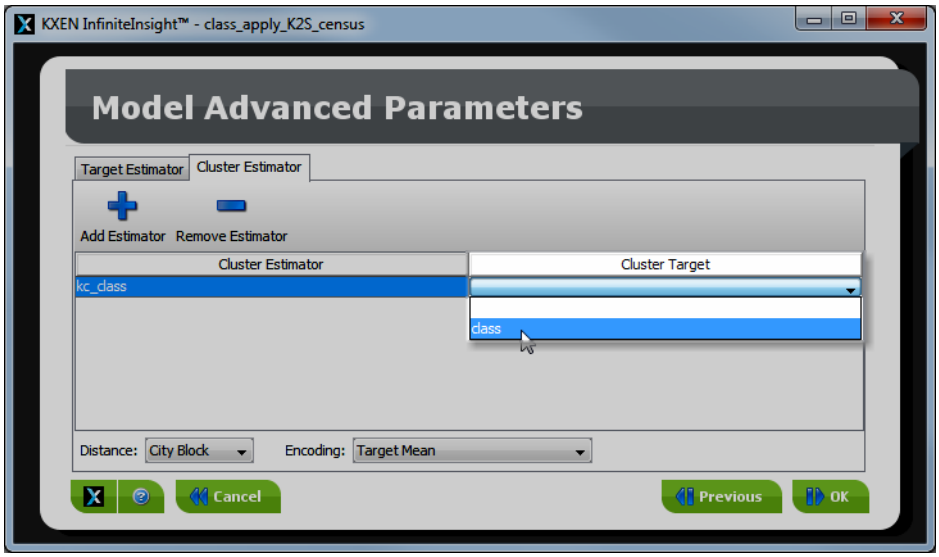
When using a data set that results from a model apply and that contains identified clusters for each record, this tab allows you to indicate the columns containing the cluster indexes. It can be useful to generate statistics, such as the size of each cluster and the residual sum of squares, on these clusters. One typical use is the comparison of clustering results obtained thanks to different tools or with different methods.

☑ To Add a Cluster Estimator

1. Select the tab *Cluster Estimator*.
2. Click the button *Add Estimator*. A variable is added to the list of estimators.
3. Click the variable name to display the list of all variables.



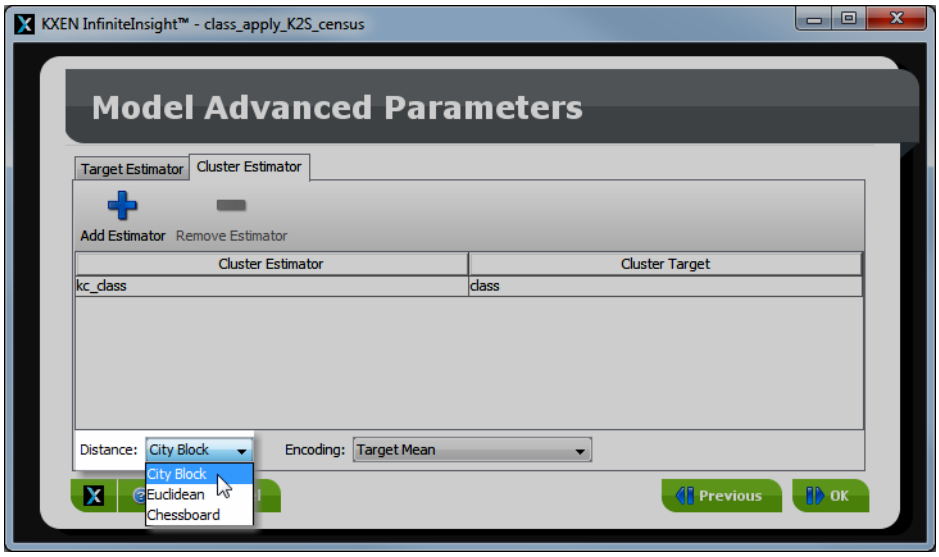
4. Select the variable containing the cluster index in the list.
5. If the clustering used to generated the cluster index was supervised, select the *Cluster Target*.



© 2014 SAP SE or an SAP affiliate company. All rights reserved-

6 Repeat steps 2 to 5, if there are more than one cluster index. For example, if the data set contains the results from two different clustering tools, so that you can compare the statistics directly on the same data set.

7 In the *Distance* list located below the estimators list, select the distance that was used to evaluate the proximity of two given individuals in the original clustering model.



The available distances are defined as follows:

- *Chessboard*: maximum of absolute differences between coordinates (LInf).
- *Euclidean*: square root of sum of square differences between coordinates (L2).
- *City Block*: sum of absolute differences between coordinates (L1).

8 In the *Encoding* list located below the estimators list, select the encoding that was used in the original clustering model. The encoding used by default when generating a clustering model with SAP InfiniteInsight® is:

- *Target Mean*, if the model is supervised,
- *Unsupervised*, otherwise.

9 Once all cluster estimators have been indicated, click the *Validate* button.

☑ To Remove a Cluster Estimator

10 Select the estimator variable to remove.

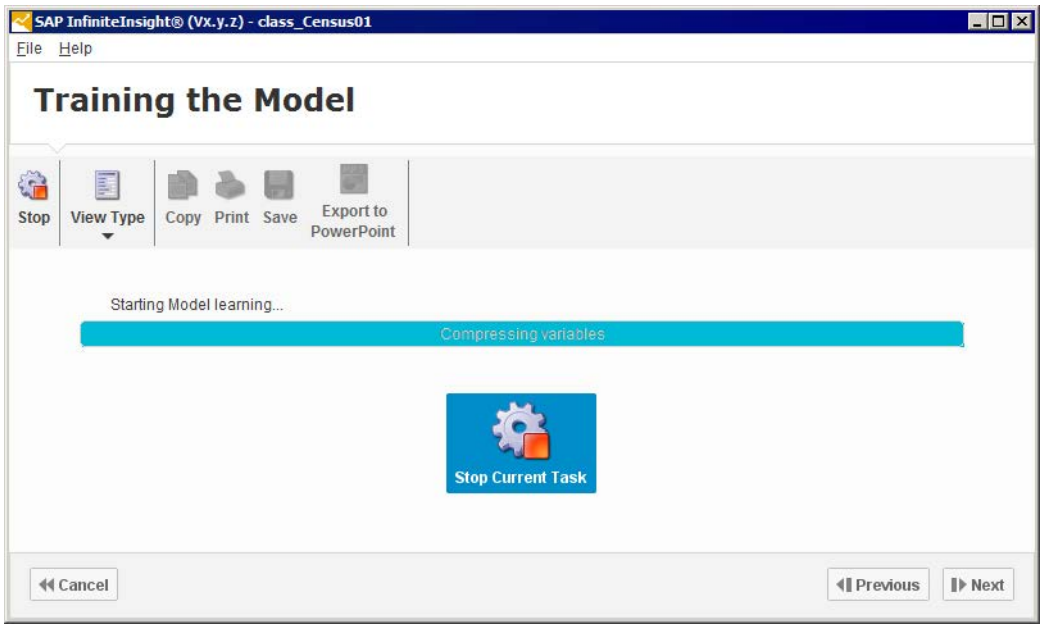
11 Click the button *Remove Estimator*.

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

6.1.3 Generating the Statistics

☑ To Generate the Statistics

1 On the screen *Summary of Modeling Parameters*, click the *Generate* button.
The screen *Training the Model* will appear. The model that will compute the statistics is being generated. A progress bar will allow you to follow the process.



2 Once the model has been generated, click the *Next* button.

6.1.4 Statistics Debriefing

The screen *Statistics Debriefing* allows you to display the statistics generated for the data set, to test a new data set for deviation and to save the statistics.

Displaying the Statistics

The *Display* section provides you with the following tools to debrief the statistics:

- the model *Summary*: displays information on the model generated such as the number of variables, the performance indicators, the target information, and so on.

  Note - If you have not used an estimator of the target, the performance indicators are not filled.

- the *Model Graphs*: depending on the type of the target, the model graph plot allows you to:
  - View the realizable profit that pertains to your business issue using the model generated when the target is nominal.
  - Compare the performance of the model generated with that of a random type model and that of a hypothetical perfect model when the target is nominal.

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

- Compare the predicted value to the actual value when the target is continuous.

Note - The model graphs are not available if you have not selected an estimator of the target beforehand.

- the *Significance of Categories*: the importance of each category of nominal or ordinal variables and the importance of each bin of the continuous variables are displayed.
- the *Statistical Reports*: SAP InfiniteInsight® provides you with a set of reports allowing you a detailed debriefing of your data set.

Running the Model

The option *Analyze Deviations* of the *Run* section allows you to load a new data set to control if any deviations exist between the new data set and the one used to build the model.

Saving the Model

The *Save* option allows you to save the statistics and to load them at a later time like any KXEN models.

6.2 From the Panel "View Data"

The panel *View Data* can be opened from various panels of SAP InfiniteInsight®, especially from any panel allowing you to select or describe a data source. It is identified by a button with a magnifier such as:

-

  or
-

This panel allows you to visualize the content of your data set and to calculate the descriptive statistics for the variables.

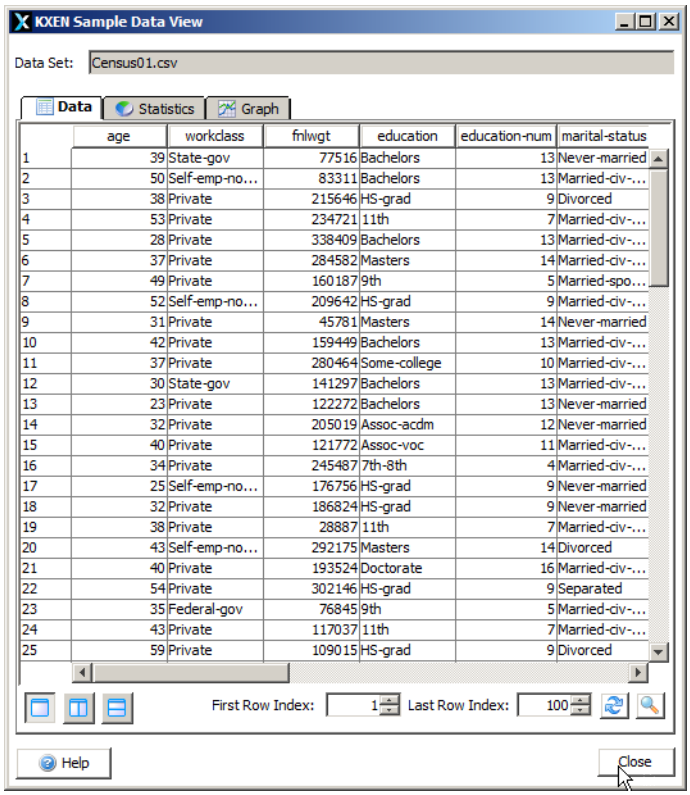
If the **data source is stored in a Oracle, Teradata or SQL Server 2005 data base**, an additional warning may be displayed, asking you to validate the execution of the query. For more information, refer to chapter Special Case: Data Stored in Databases (on page 5).

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

6.2.1 Computing the Statistics

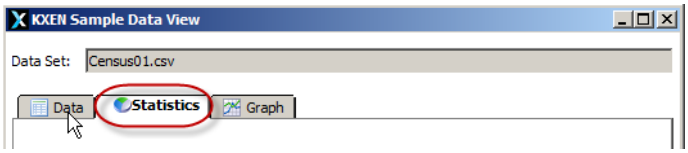
☑ To Compute the Statistics

1 Click the *View Data* button. The panel *KXEN Sample Data View* opens.



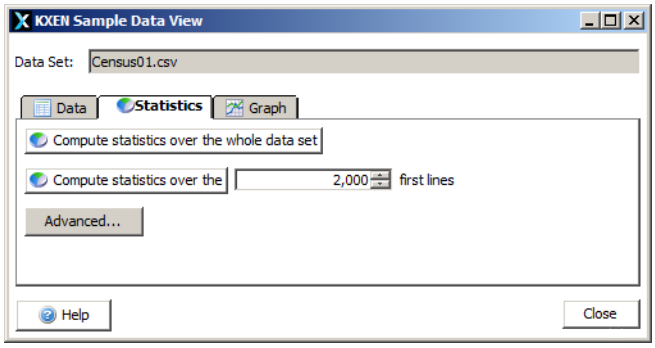
	age	workclass	fnlwgt	education	education-num	marital-status
1	39	State-gov	77516	Bachelors	13	Never-married
2	50	Self-emp-no...	83311	Bachelors	13	Married-civ-...
3	38	Private	215646	HS-grad	9	Divorced
4	53	Private	234721	11th	7	Married-civ-...
5	28	Private	338409	Bachelors	13	Married-civ-...
6	37	Private	284582	Masters	14	Married-civ-...
7	49	Private	160187	9th	5	Married-spo...
8	52	Self-emp-no...	209642	HS-grad	9	Married-civ-...
9	31	Private	45781	Masters	14	Never-married
10	42	Private	159449	Bachelors	13	Married-civ-...
11	37	Private	280464	Some-college	10	Married-civ-...
12	30	State-gov	141297	Bachelors	13	Married-civ-...
13	23	Private	122272	Bachelors	13	Never-married
14	32	Private	205019	Assoc-acdm	12	Never-married
15	40	Private	121772	Assoc-voc	11	Married-civ-...
16	34	Private	245487	7th-8th	4	Married-civ-...
17	25	Self-emp-no...	176756	HS-grad	9	Never-married
18	32	Private	186824	HS-grad	9	Never-married
19	38	Private	28887	11th	7	Married-civ-...
20	43	Self-emp-no...	292175	Masters	14	Divorced
21	40	Private	193524	Doctorate	16	Married-civ-...
22	54	Private	302146	HS-grad	9	Separated
23	35	Federal-gov	76845	9th	5	Married-civ-...
24	43	Private	117037	11th	7	Married-civ-...
25	59	Private	109015	HS-grad	9	Divorced

2 Select the *Statistics* tab.



3 Select on which part of the data set you want to compute the statistics. You can calculate the statistics for:

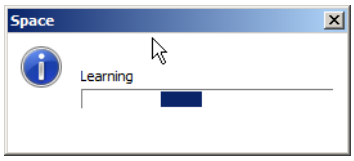
- the whole data set (option *Compute statistics over the whole data set*),
- the first *n* lines of the data set (option *Compute statistics over the n first lines*),
- a sub-sample of the data set (*Advanced* option).



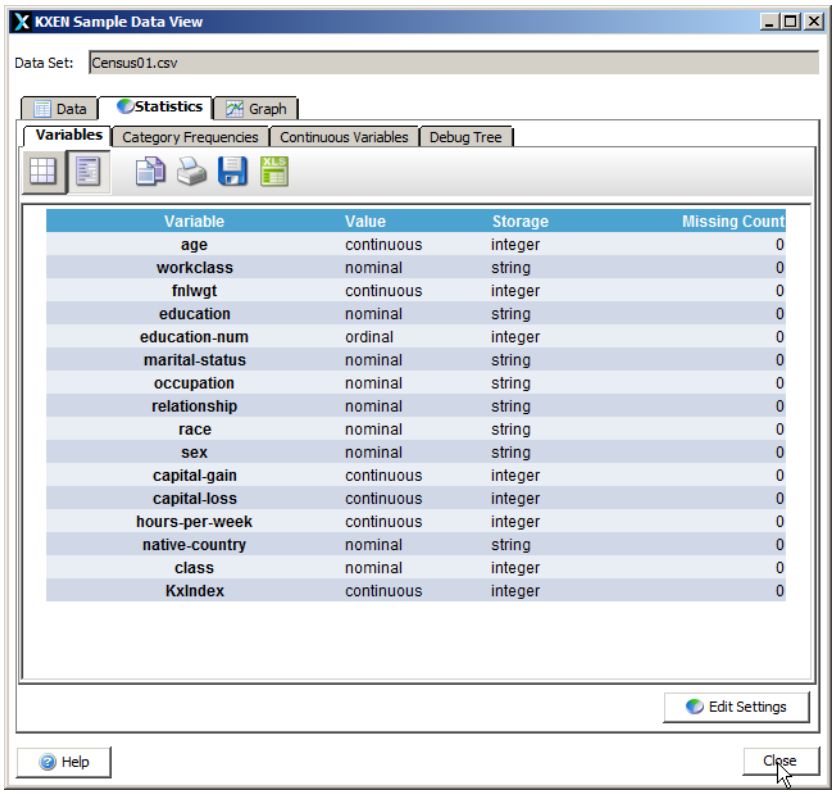
© 2014 SAP SE or an SAP affiliate company. All rights reserved-

The time needed to compute the statistics depends on the quantity of data to analyze, so **the larger the number of lines, the longer the analysis.**

When you click on the selected option, the statistics generation starts. A progress bar is displayed allowing you to follow the process.



When the analysis is over, the statistics are displayed.



Variable	Value	Storage	Missing Count
age	continuous	integer	0
workclass	nominal	string	0
fnlwgt	continuous	integer	0
education	nominal	string	0
education-num	ordinal	integer	0
marital-status	nominal	string	0
occupation	nominal	string	0
relationship	nominal	string	0
race	nominal	string	0
sex	nominal	string	0
capital-gain	continuous	integer	0
capital-loss	continuous	integer	0
hours-per-week	continuous	integer	0
native-country	nominal	string	0
class	nominal	integer	0
KxIndex	continuous	integer	0

6.2.2 Viewing the Statistics

Once the statistics have been computed, the following tabs are displayed:

- *Variables*, which displays a list of the variables found in the data set with their value, their storage value, the number of missing value for each variable and their role during the modeling.
- *Category Frequencies*, which displays for each variable the complete set of its categories, their frequency and the number of occurrences of each category (*Weighted Count*).
- *Continuous Variables*, which displays for each continuous variable the minimum, the maximum, the mean and the standard deviation.

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

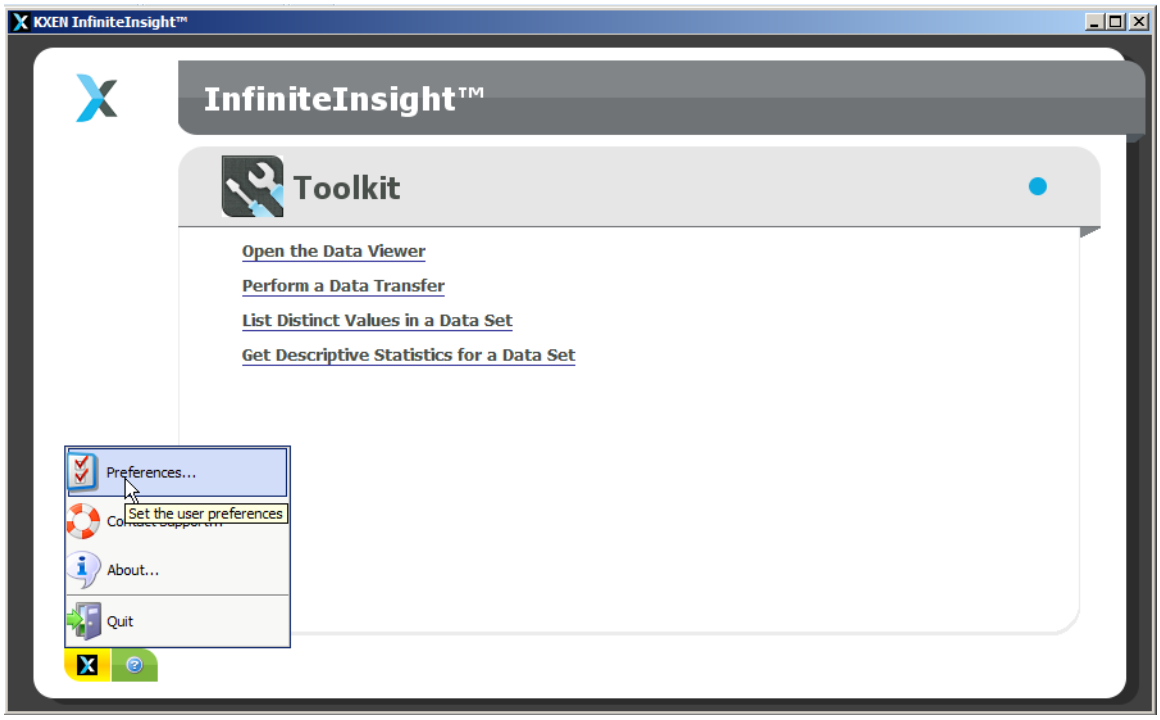
6.3 Geo-location

Pre-requisites:

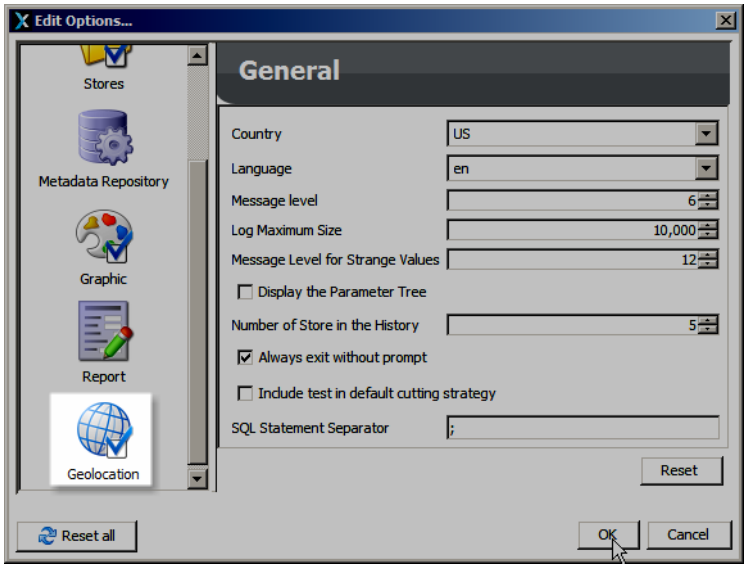
- Install GoogleEarth
- Add Preferences in the Geolocation section.

☑ To Add Preferences in the Geolocation Section

1 Launch the SAP InfiniteInsight® Preferences panel.

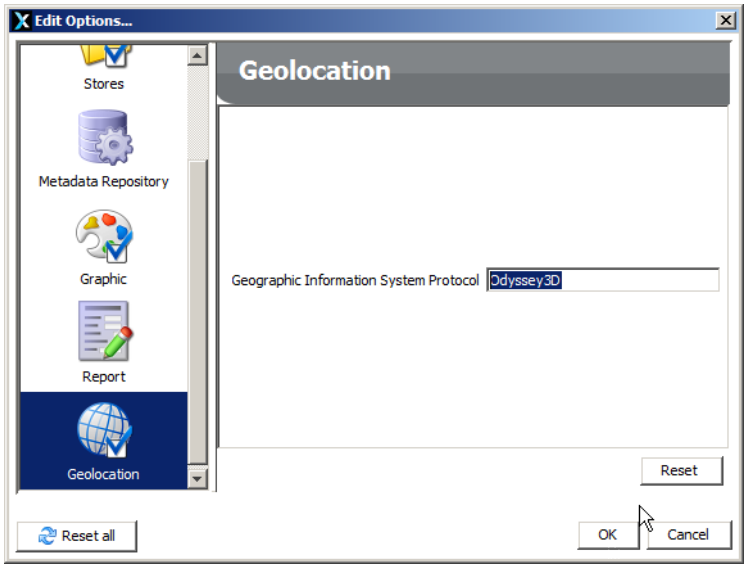


The Edit Options screen appears.



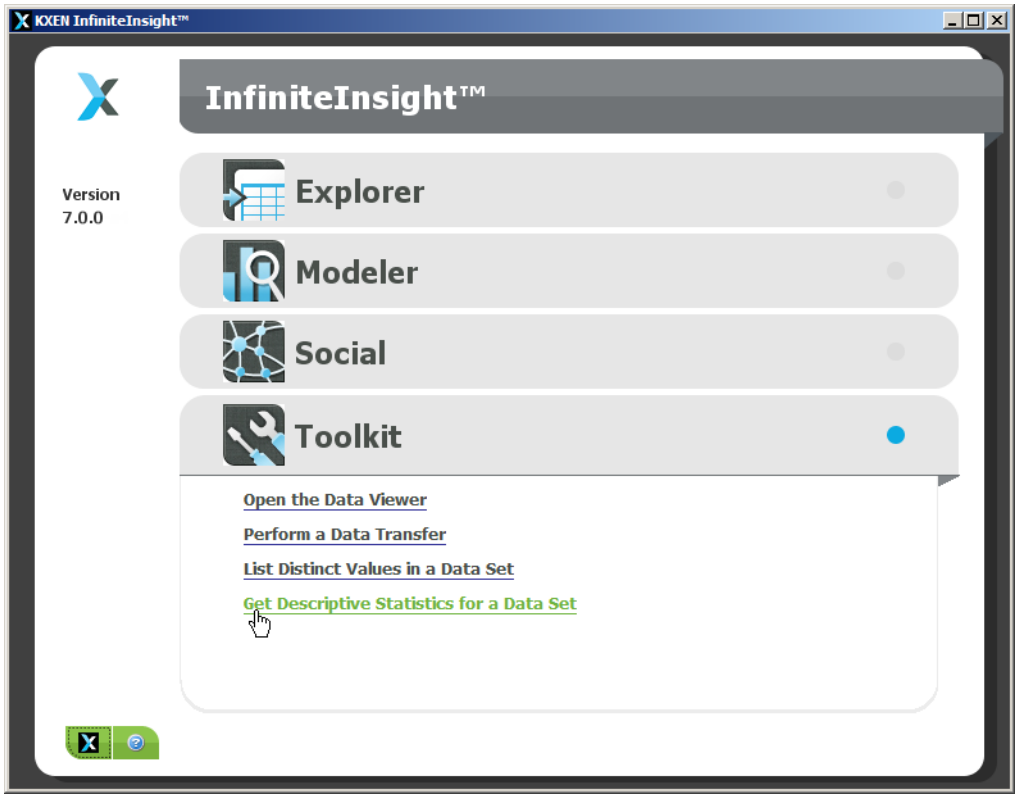
© 2014 SAP SE or an SAP affiliate company. All rights reserved-

2 Go to the section Geolocation.
3 Enter a Geographic Information System Protocol
4 Click OK to confirm.



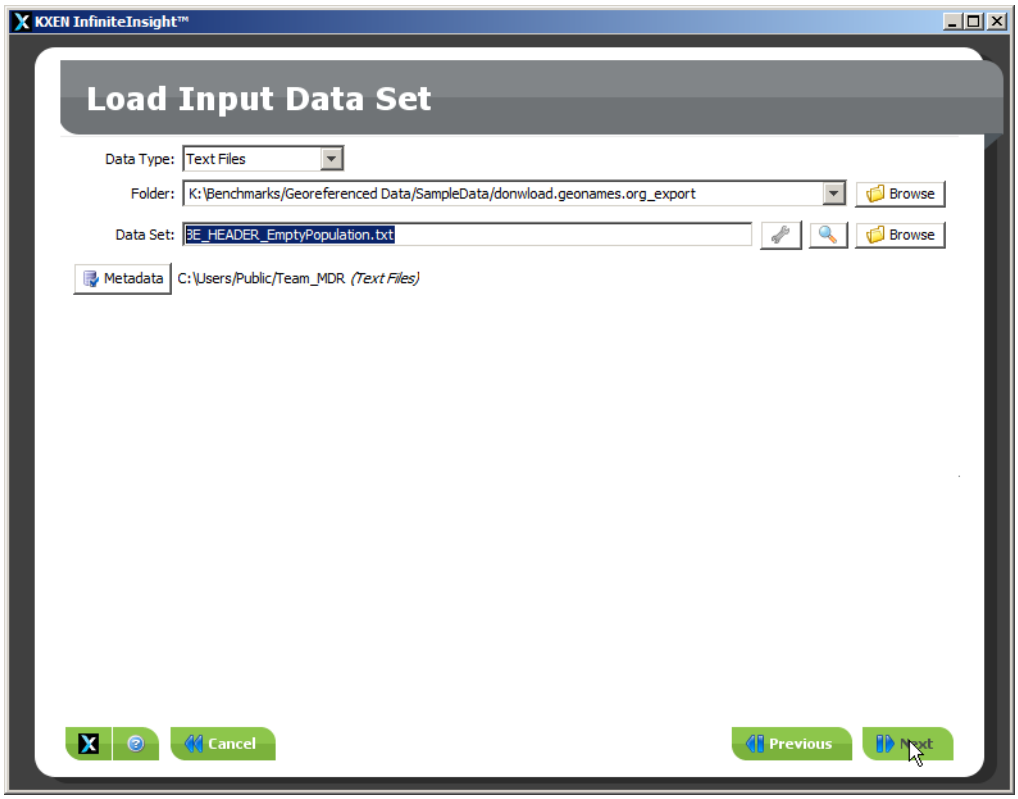
6.3.1 Using Geolocation Feature

1 InfiniteInsight Start Panel > Data Toolkit section > **Get Descriptive Statistics for a Data Set.**



© 2014 SAP SE or an SAP affiliate company. All rights reserved-

The screen Load Input Data Set is displayed

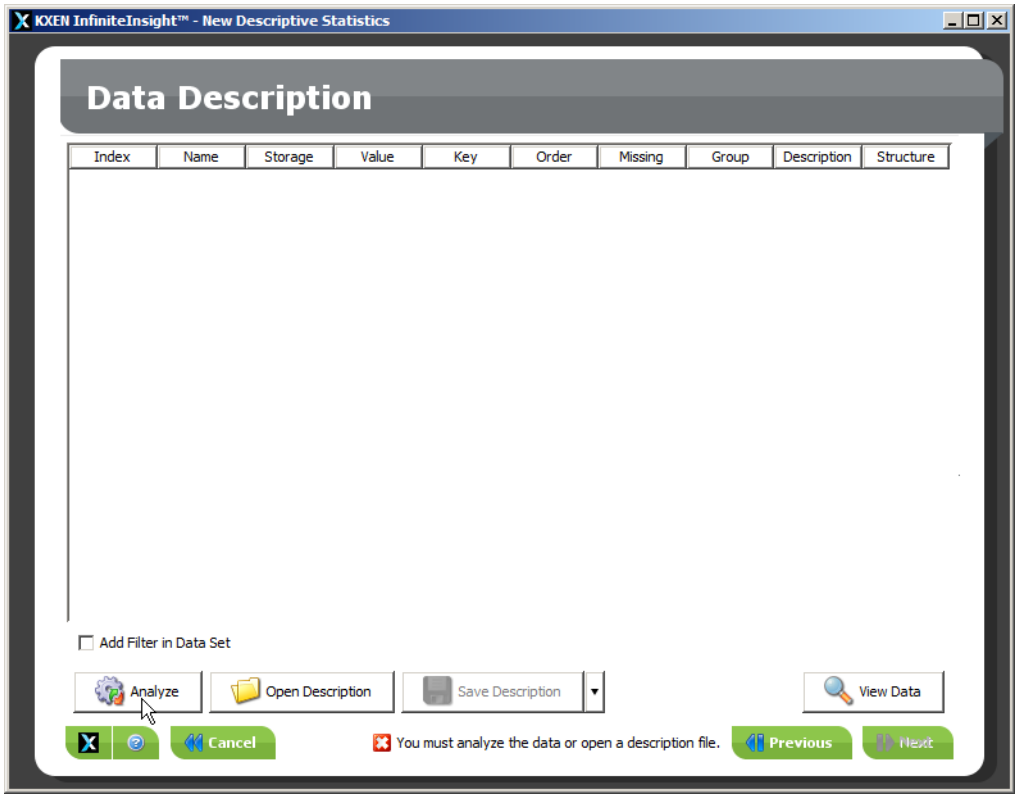


2 Select a Data Type, a Folder and a Data Set.
For the example:
K:\Benchmarks/Georeferenced Data/SampleData/donwload.geonames.org_export
and select file BE_HEADER_EmptyPopulation.txt (where there is a target variable).

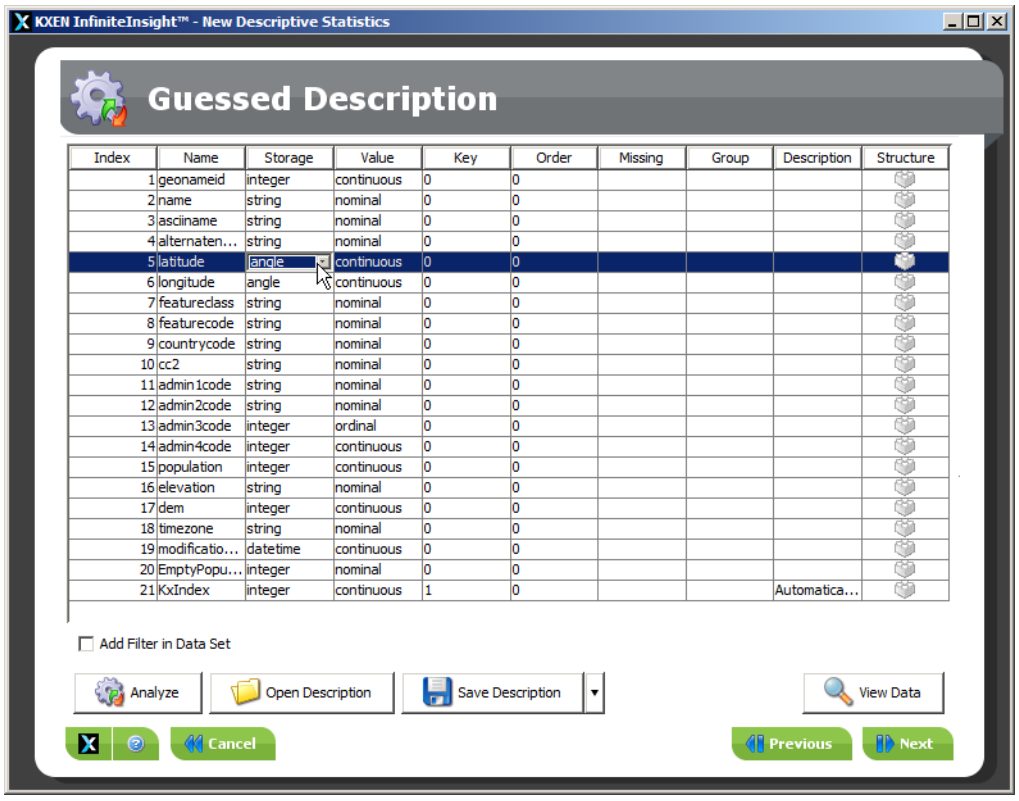
© 2014 SAP SE or an SAP affiliate company. All rights reserved-

3 Click *Next*.

4 In the panel **Data Description**, click *Analyze*.

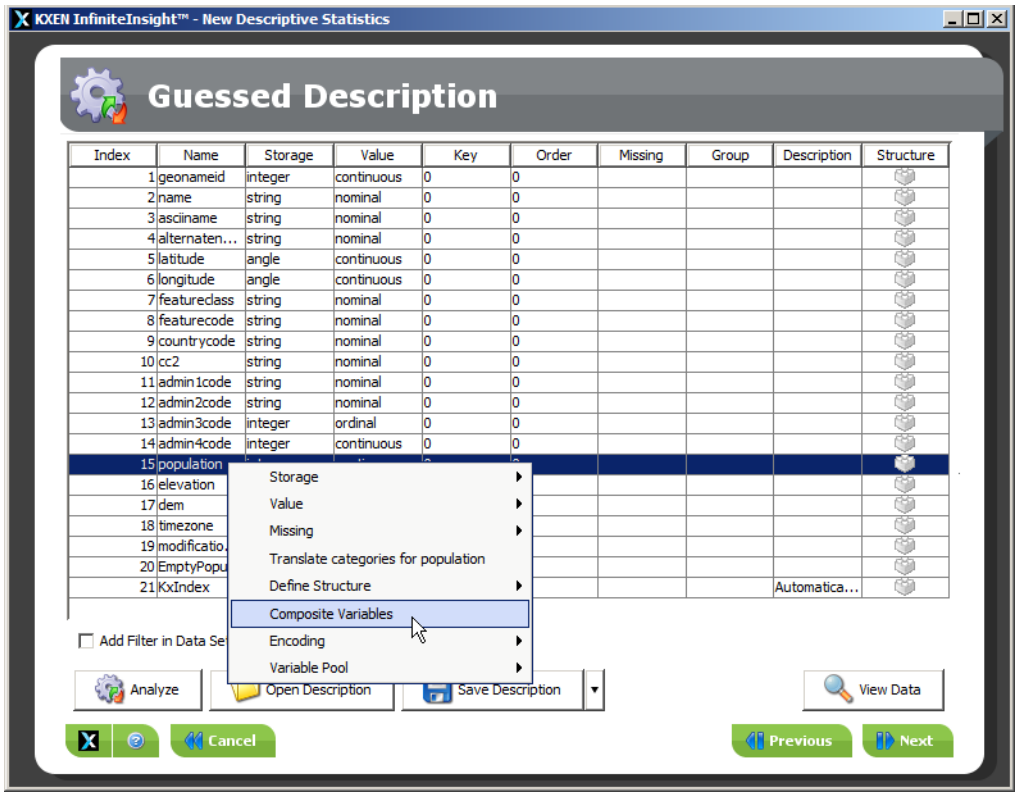


The described data appears.


KXEN InfiniteInsight™ - New Descriptive Statistics
Guessed Description
Index Name Storage Value Key Order Missing Group Description Structure
1 geonameid integer continuous 0 0
2 name string nominal 0 0
3 asciiname string nominal 0 0
4 alternaten... string nominal 0 0
5 latitude angle continuous 0 0
6 longitude angle continuous 0 0
7 featureclass string nominal 0 0
8 featurecode string nominal 0 0
9 countrycode string nominal 0 0
10 cc2 string nominal 0 0
11 admin1code string nominal 0 0
12 admin2code string nominal 0 0
13 admin3code integer ordinal 0 0
14 admin4code integer continuous 0 0
15 population integer continuous 0 0
16 elevation string nominal 0 0
17 dem integer continuous 0 0
18 timezone string nominal 0 0
19 modificatio... datetime continuous 0 0
20 EmptyPopu... integer nominal 0 0
21 KxIndex integer continuous 1 0 Automatica...
Add Filter in Data Set
Analyze
Open Description
Save Description
View Data
Cancel
Previous
Next

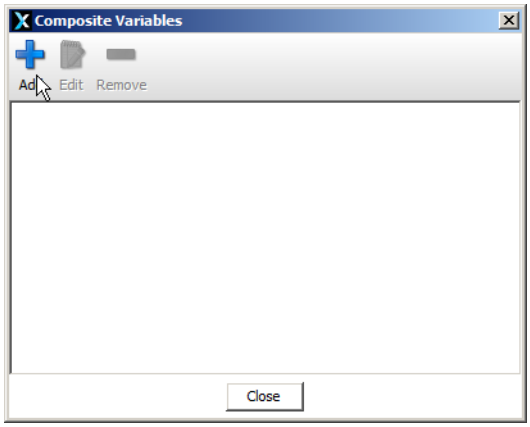
© 2014 SAP SE or an SAP affiliate company. All rights reserved-

5 Select a variable, click right, choose Composite Variables.



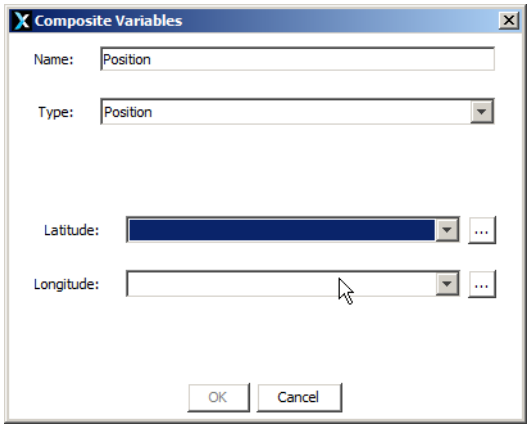
The panel Composite Variables appears.

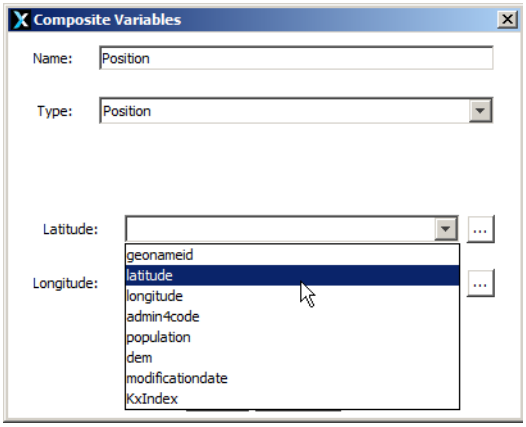
6 To Add a Composite Variable, click + *Add*.

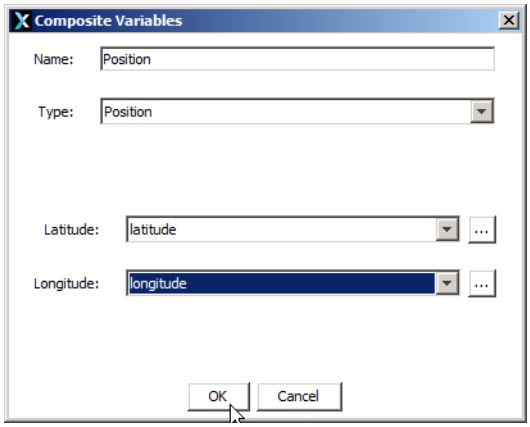


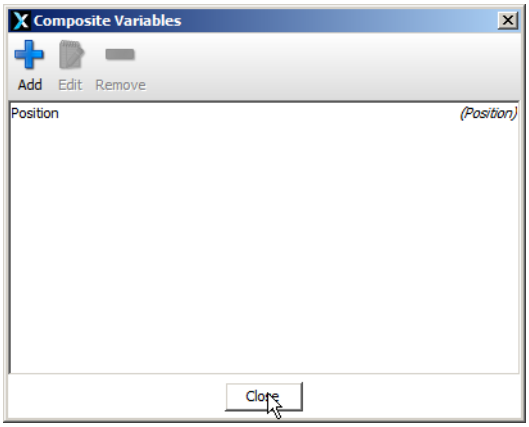
© 2014 SAP SE or an SAP affiliate company. All rights reserved-

7 Enter a Name, a Type, a Latitude, a Longitude.









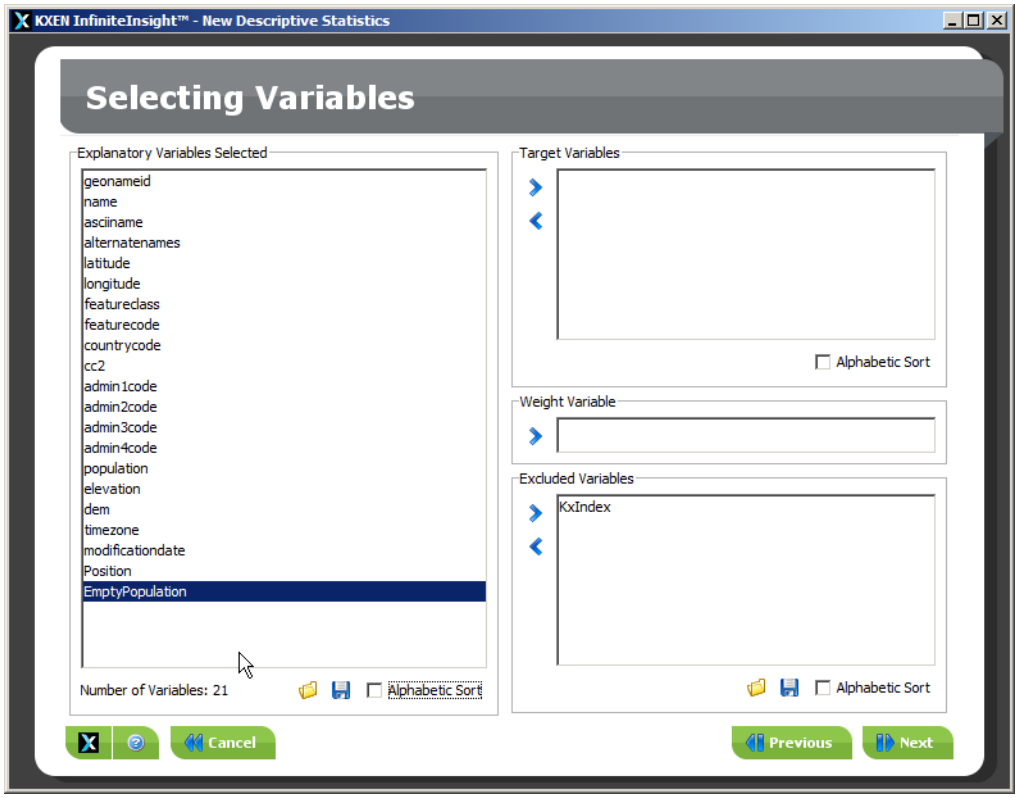
© 2014 SAP SE or an SAP affiliate company. All rights reserved-

8 Click OK. The Composite Varible newly created is displayed.

9 Click Close.

10 Back to the Description panel, click Next.

11 For this scenario with a Target Variable, select the variable EmptyPopulation.

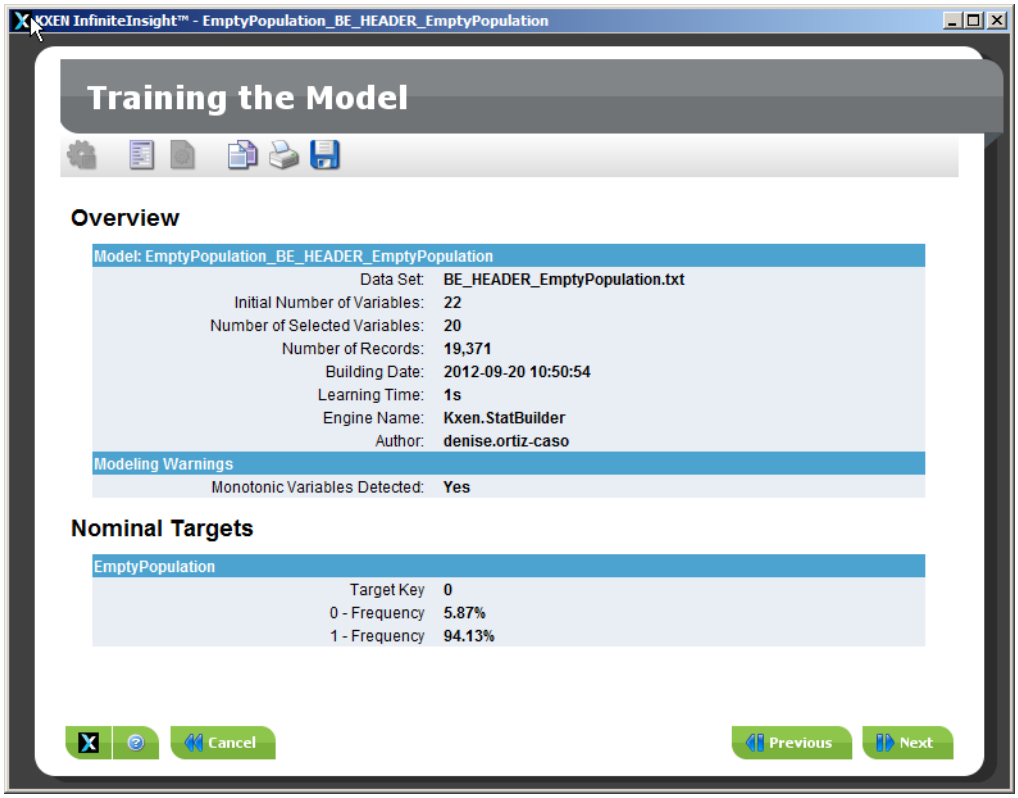


© 2014 SAP SE or an SAP affiliate company. All rights reserved-

12 Click Next.

13 Create the model by clicking Generate. The Training the Model screen appears.

14 Click *Next*.



The panel Statistics Debriefing is displayed.

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

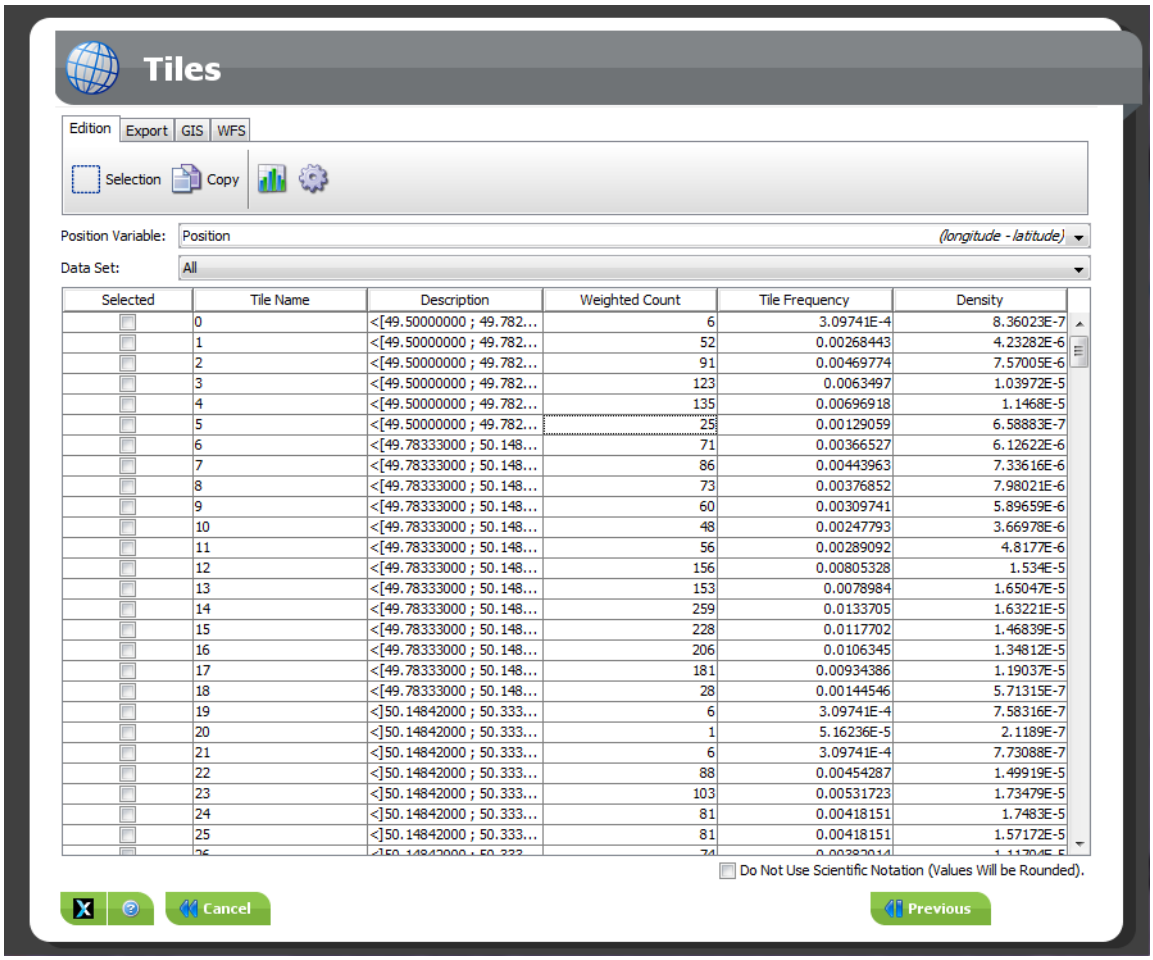
15 Go to section Tiles.

The Tiles screen appears. It is composed of four tabs:

- Edition
- Export
- GIS (Geographic Information System)

© 2014 SAP SE or an SAP affiliate company. All rights reserved-

- WFS (Web Feature Server)

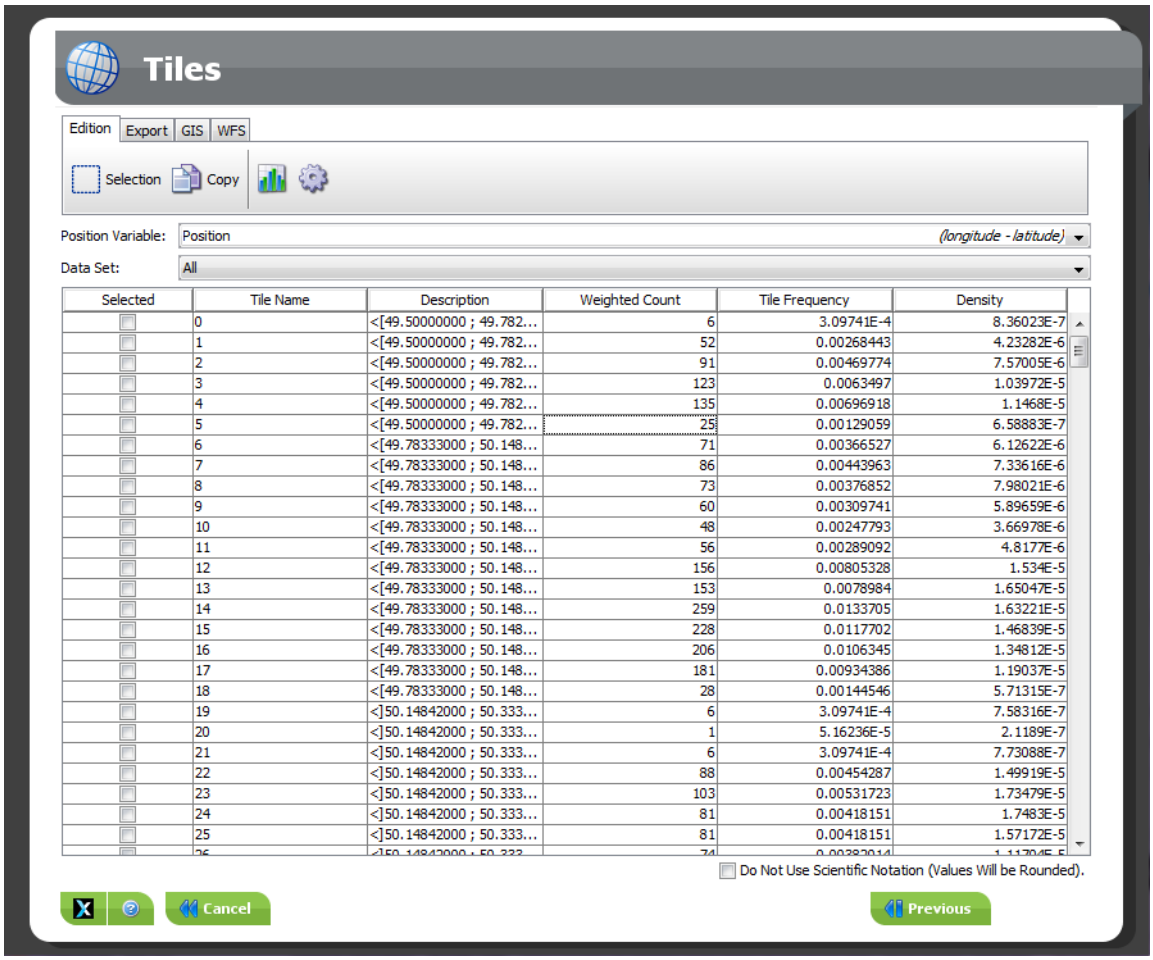


© 2014 SAP SE or an SAP affiliate company. All rights reserved-

Edition

The Edition tab shows the Postion Variable, the Data Set used for the Geolocation and the Target.

It also offers the options Selection, Copy, Graph and Settings.



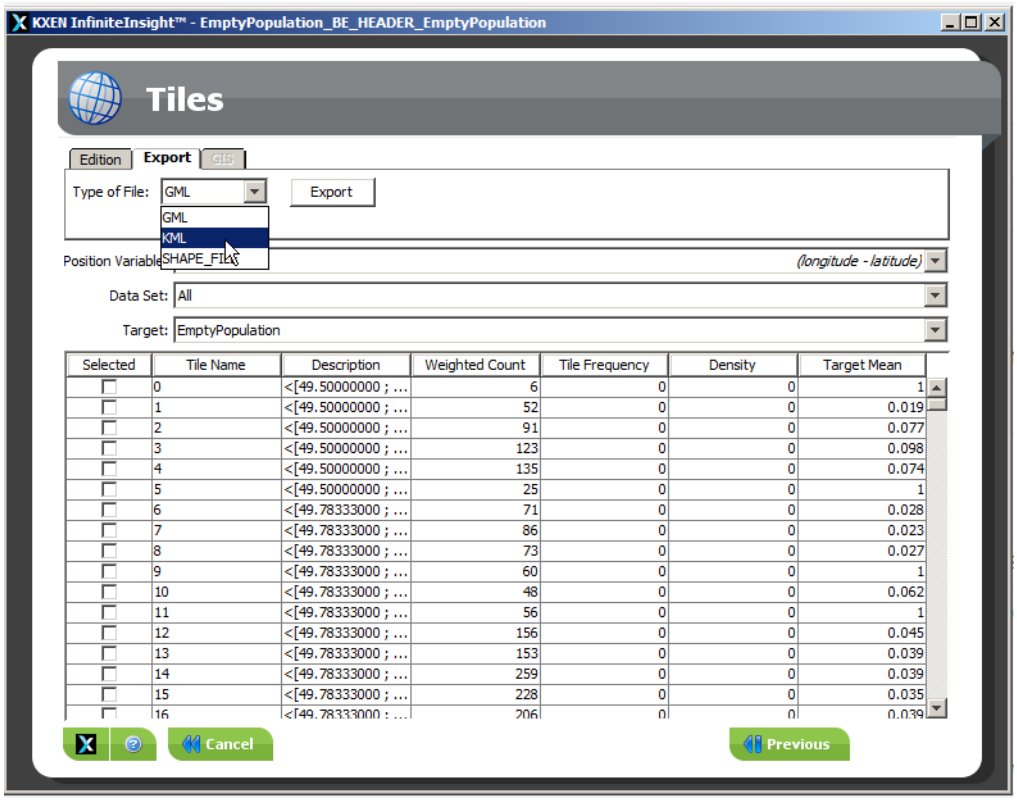
© 2014 SAP SE or an SAP affiliate company. All rights reserved-

Export

The export tab enables you to export the Selected or All the Tiles of your model and to open the output file with GoogleEarth software.

☑ To Export the Selected or All the Tiles

1 Select File Type KML.

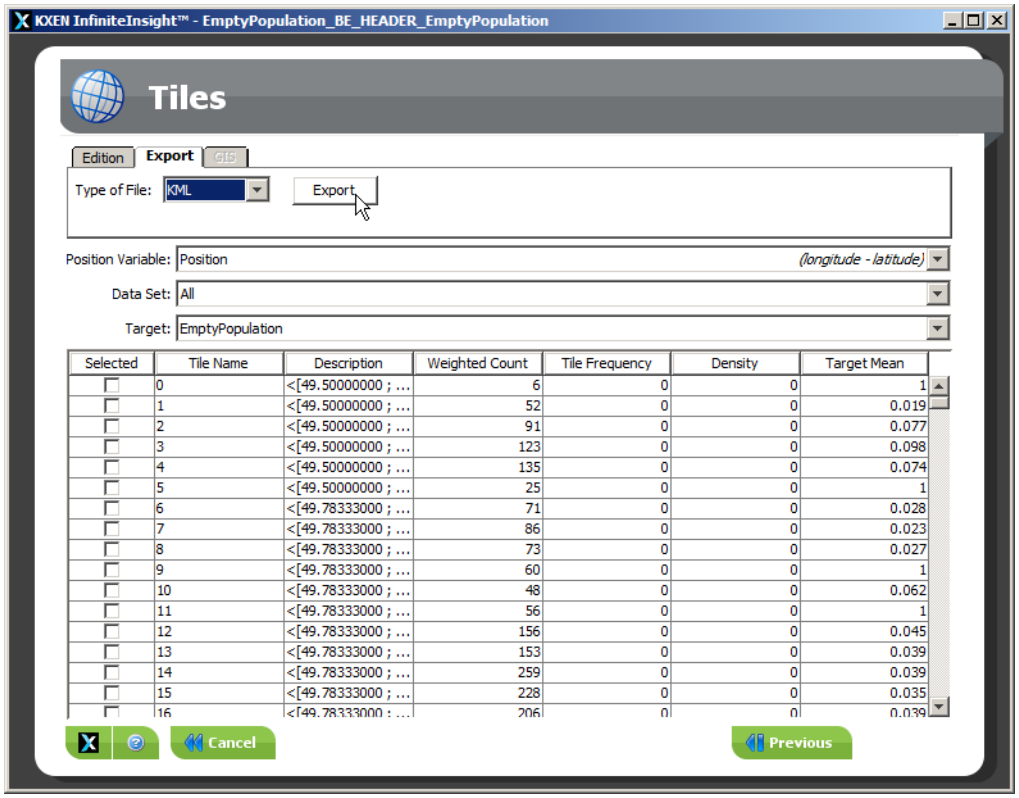


© 2014 SAP SE or an SAP affiliate company. All rights reserved-

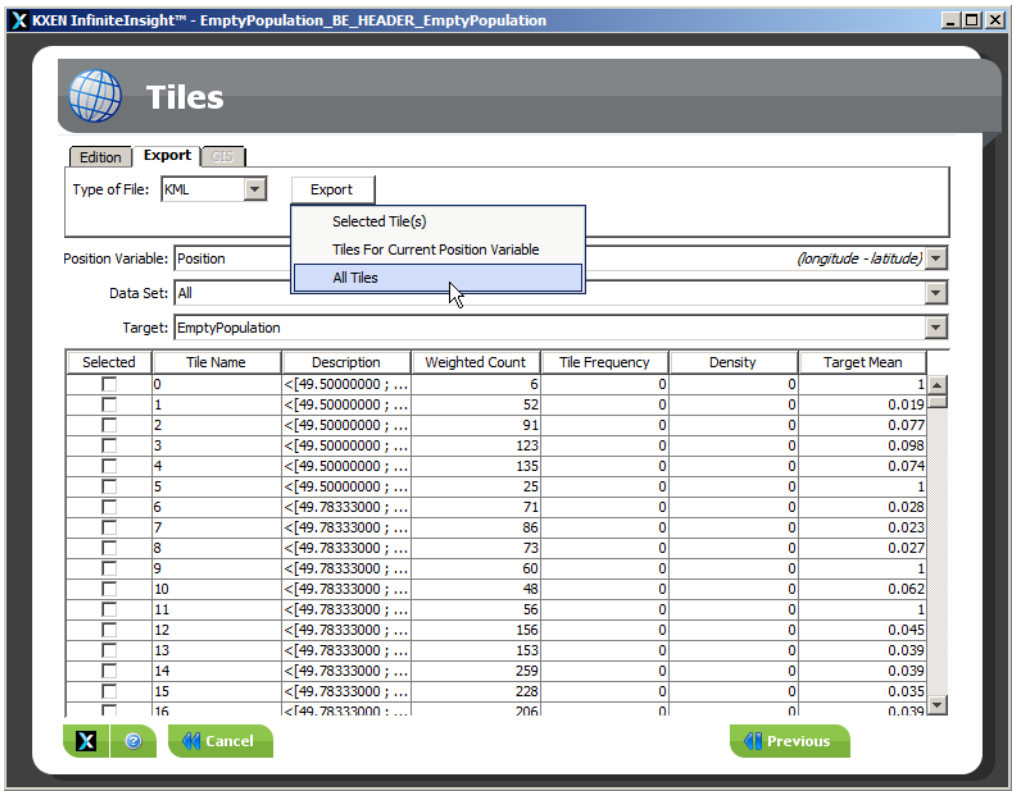
2 Check the fields Position Variables, Data Set and Target.

3 Select the Tiles you need to export or keep them all.

4 Click Export.

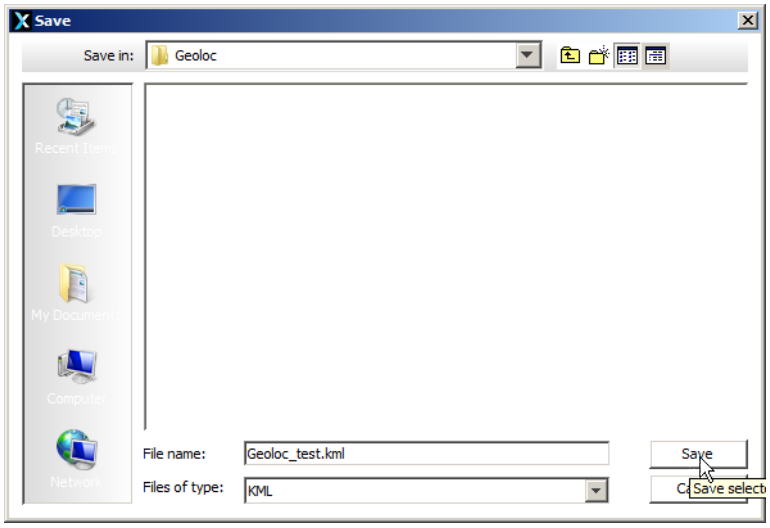


5 Precise which tiles you need to export.



© 2014 SAP SE or an SAP affiliate company. All rights reserved-

6 Save the file.

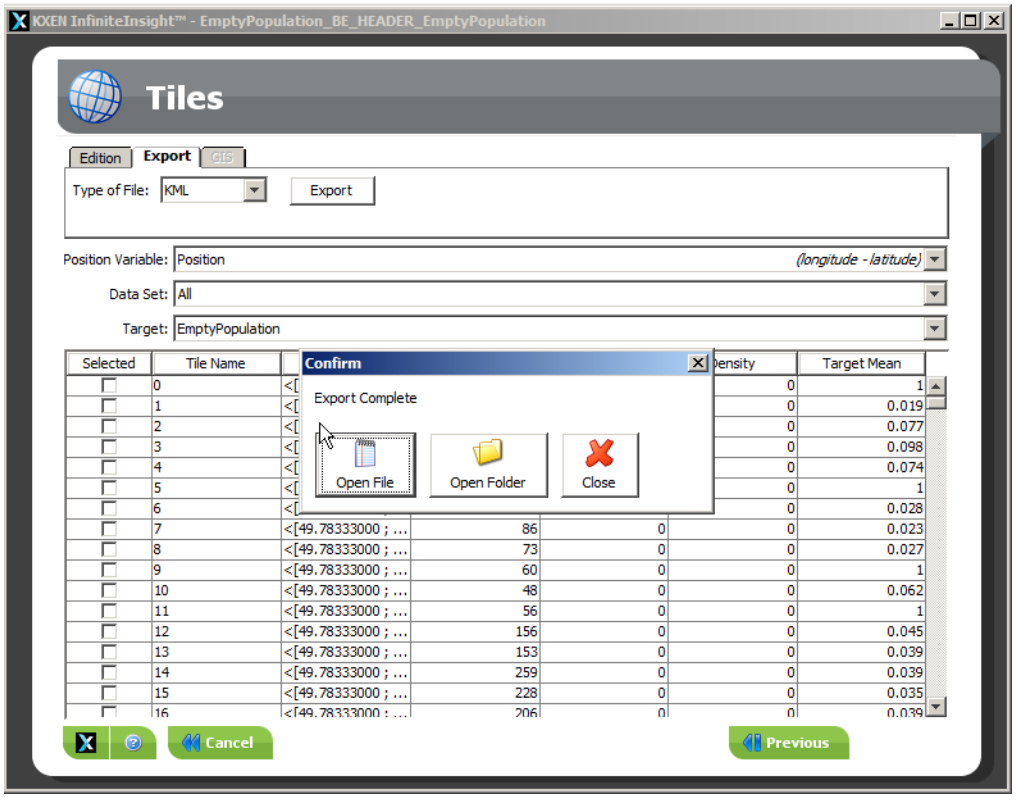


7 A pop window indicates that the export has been completed and that the file can be opened.

Note -

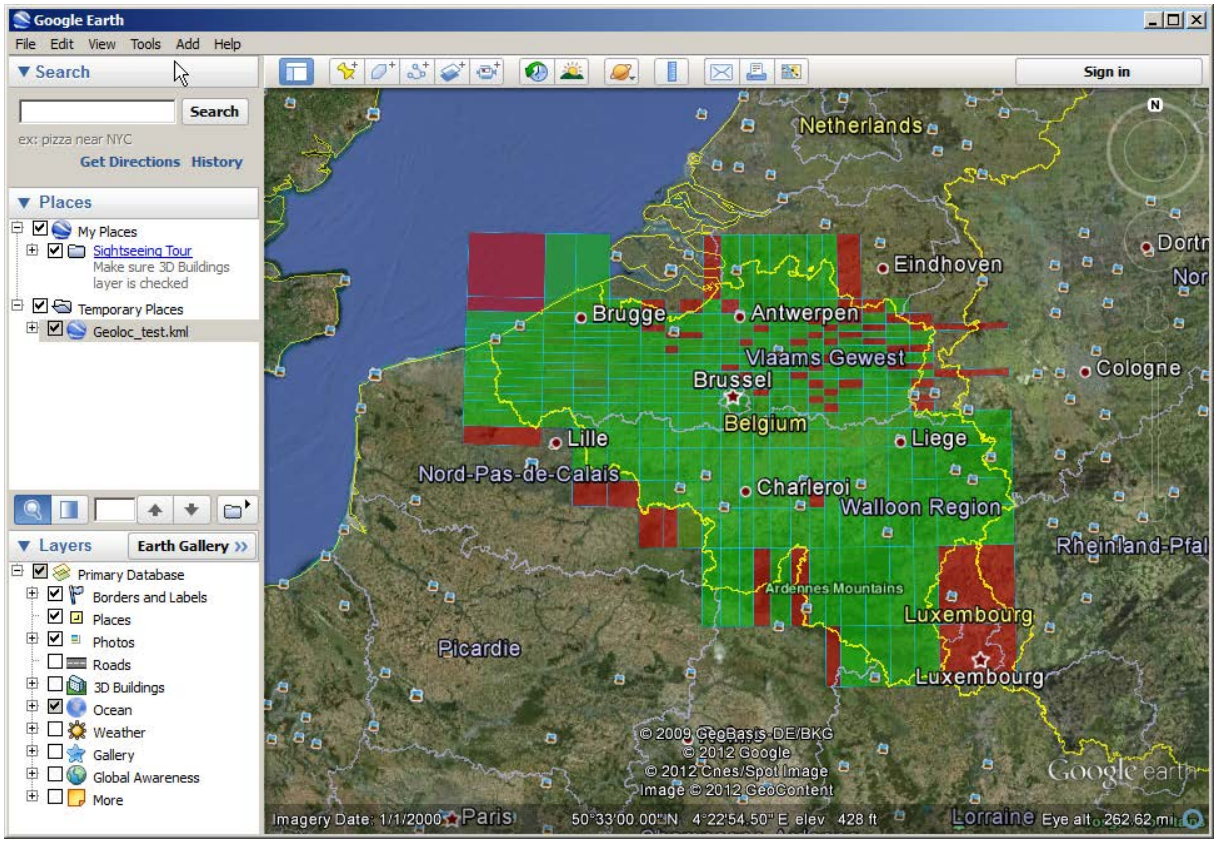
For the first export: click Close. Go to the folder where the file was saved and open it manually with GoogleEarth software. Check the box *Always use the selected program to open this kind of file* so that you will be able to open directly your file from InfiniteInsight interface.

8 Open the file



The Tiles appear in GoogleEarth.

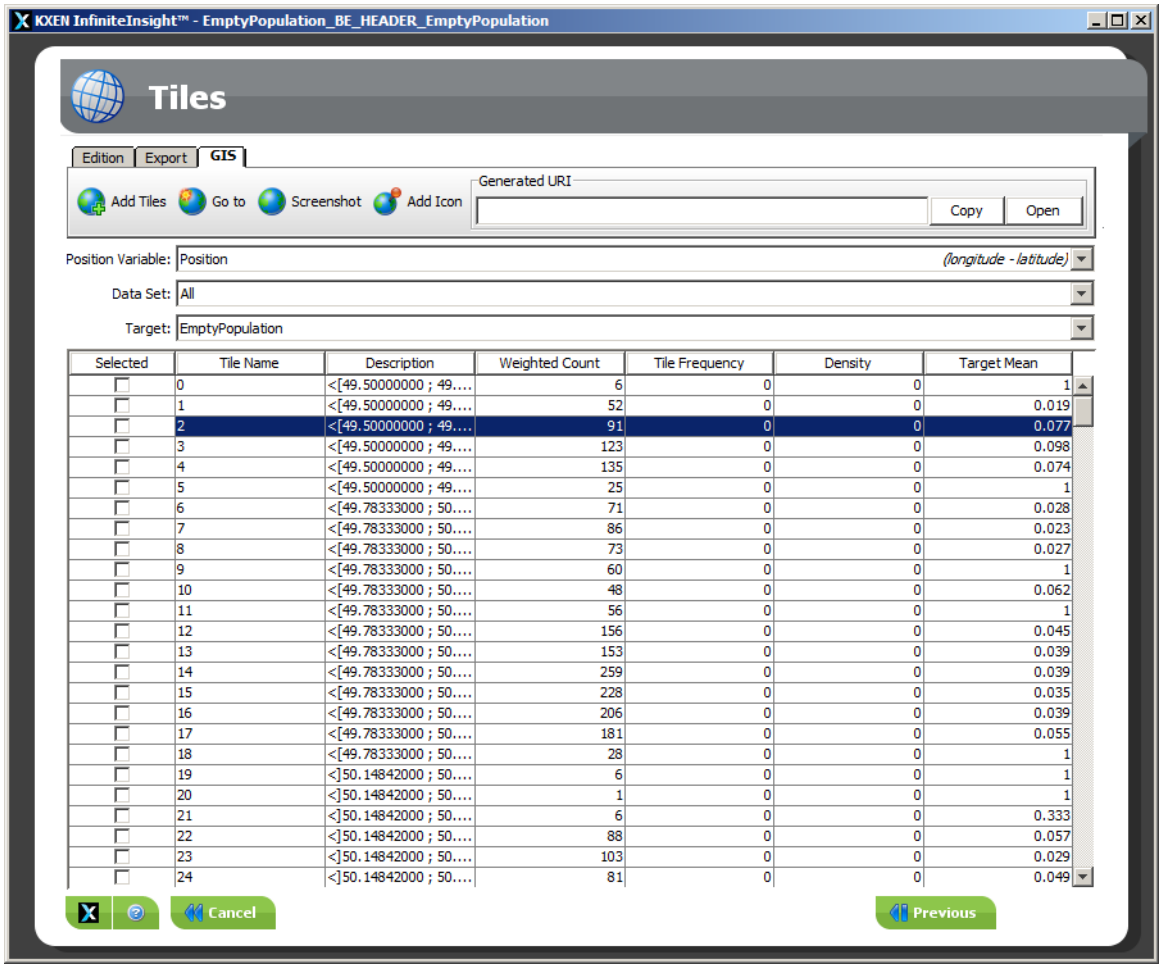
© 2014 SAP SE or an SAP affiliate company. All rights reserved-



© 2014 SAP SE or an SAP affiliate company. All rights reserved-

Geographic Information System (GIS)

The GIS feature is an interface with another Geo Software. In this scenario, the software is Odyssey3D.

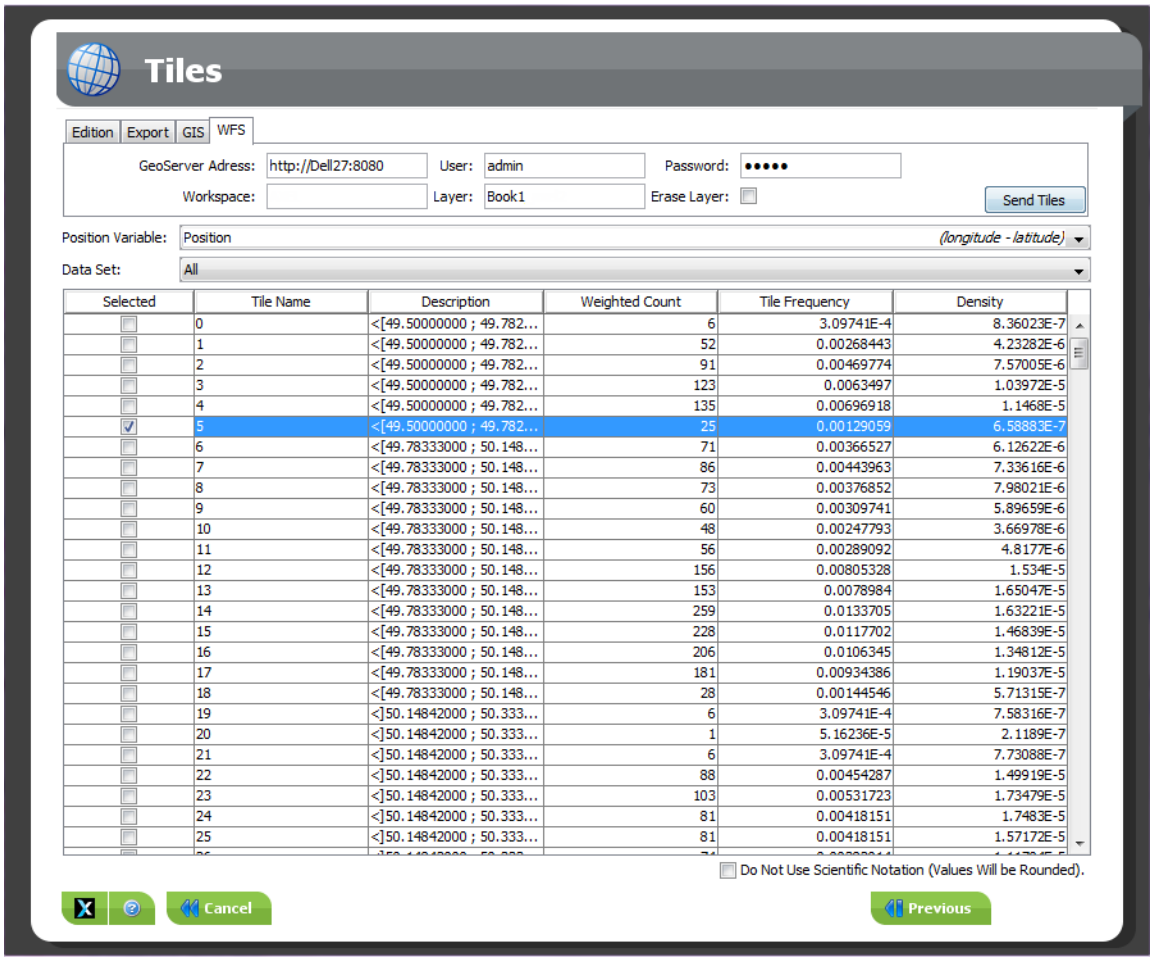


Selected	Tile Name	Description	Weighted Count	Tile Frequency	Density	Target Mean
	0	<[49.50000000 ; 49....	6	0	0	1
	1	<[49.50000000 ; 49....	52	0	0	0.019
	2	<[49.50000000 ; 49....	91	0	0	0.077
	3	<[49.50000000 ; 49....	123	0	0	0.098
	4	<[49.50000000 ; 49....	135	0	0	0.074
	5	<[49.50000000 ; 49....	25	0	0	1
	6	<[49.78333000 ; 50....	71	0	0	0.028
	7	<[49.78333000 ; 50....	86	0	0	0.023
	8	<[49.78333000 ; 50....	73	0	0	0.027
	9	<[49.78333000 ; 50....	60	0	0	1
	10	<[49.78333000 ; 50....	48	0	0	0.062
	11	<[49.78333000 ; 50....	56	0	0	1
	12	<[49.78333000 ; 50....	156	0	0	0.045
	13	<[49.78333000 ; 50....	153	0	0	0.039
	14	<[49.78333000 ; 50....	259	0	0	0.039
	15	<[49.78333000 ; 50....	228	0	0	0.035
	16	<[49.78333000 ; 50....	206	0	0	0.039
	17	<[49.78333000 ; 50....	181	0	0	0.055
	18	<[49.78333000 ; 50....	28	0	0	1
	19	<]50.14842000 ; 50....	6	0	0	1
	20	<]50.14842000 ; 50....	1	0	0	1
	21	<]50.14842000 ; 50....	6	0	0	0.333
	22	<]50.14842000 ; 50....	88	0	0	0.057
	23	<]50.14842000 ; 50....	103	0	0	0.029
	24	<]50.14842000 ; 50....	81	0	0	0.049

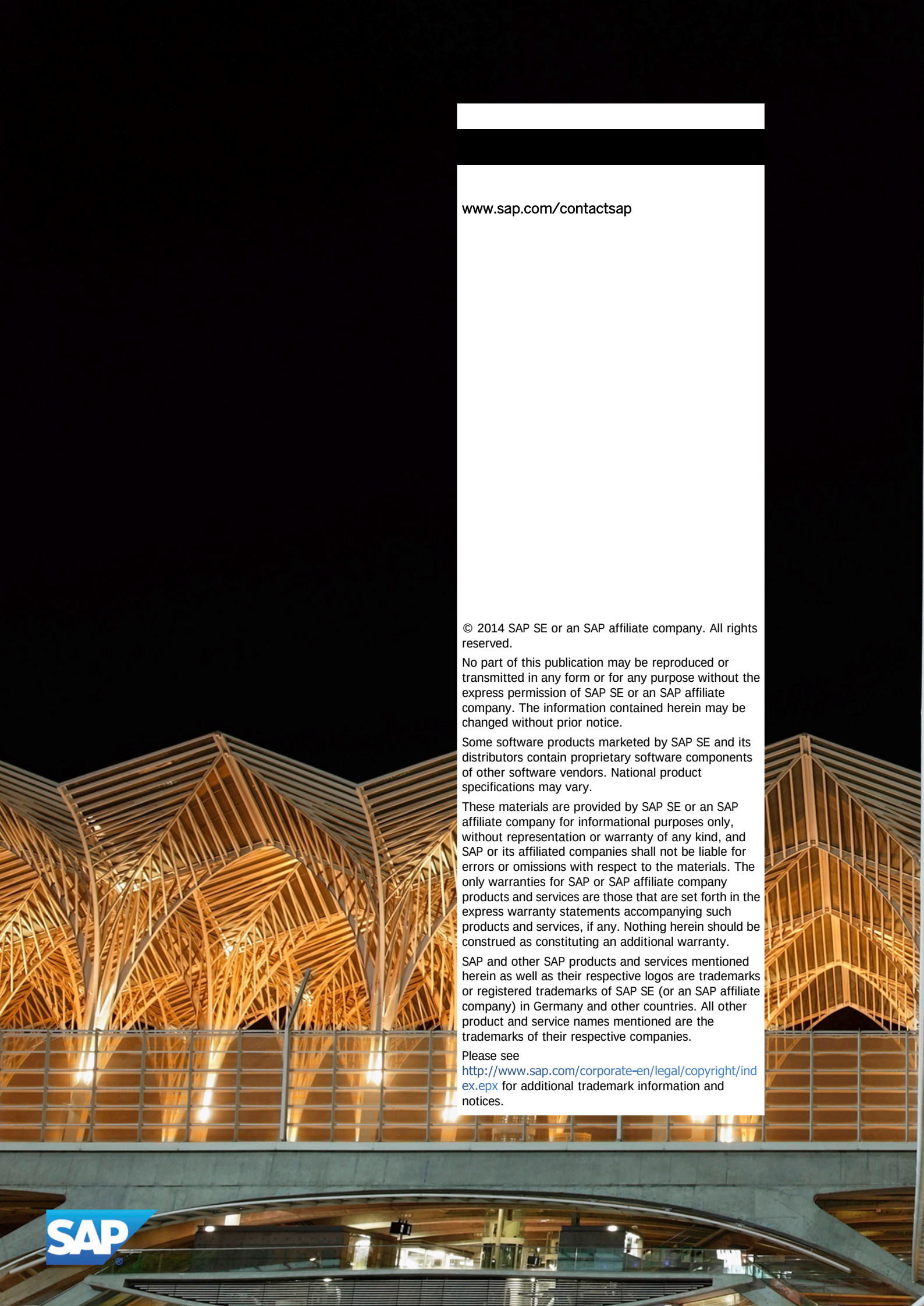
CUSTOMER
© 2014 SAP SE or an SAP affiliate company. All rights reserved-

Web Feature Server (WFS)

The WFS feature is used to load the selected tiles from a dedicated geoserver.



CUSTOMER
© 2014 SAP SE or an SAP affiliate company. All rights reserved-



www.sap.com/contactsap

© 2014 SAP SE or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company. The information contained herein may be changed without prior notice.

Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. All other product and service names mentioned are the trademarks of their respective companies.

Please see http://www.sap.com/corporate-en/legal/copyright/index.epx for additional trademark information and notices.